S. ADM.
BRARY

SEP 16 1949

Journal of the AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER 1949

The Monte Carlo Method — *Nicholas Metropolis and S. Ulam* 335

Applications of Some Significance Tests for the Median Which Are Valid Under Very General Conditions — *John E. Walsh* 342

A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis — *Guy H. Orcutt and Donald Cochrane* 356

Control of a General Census by Means of an Area Sampling Method — *Gabriel Chevry* 373

A Procedure for Objective Respondent Selection Within the Household — *Leslie Kish* 380

Beneficiary Statistics Under the Old-Age and Survivors Insurance Program and Some Possible Demographic Studies Based on These Data — *Robert J. Myers* 388

By-Product Data and Forecasting in Unemployment Insurance — *Nathan Morrison* 397

Statistical Requirements for Economic Mobilization — *Ralph J. Watkins* 406

The War Production Board's Statistical Reporting Experience, V and VI — *David Novick and George A. Steiner* 413

BOOK REVIEWS by T. A. Bancroft, Alfred Cahen, S. Lee Crump, Henry K. Evans, Lila F. Knudsen, Howard Levene, Charles M. Mottley, E. J. G. Pitman, and Herbert Robbins

LETTERS ABOUT BOOKS *by F. Stuart Chapin*

VOLUME 44 NUMBER 247

CHAPTER SECRETARIES

ALBANY	Murray Dorkin, *40 Glenwood Street, Albany, New York*
ASSOCIACIÓN CUBANA DE ESTADISTICA	Hugo Vivo, *Calle 28 No. 270, Apt. B, entre 21y 23, Vedado, Habana, Cuba*
AUSTIN	John R. Stockton, *113 Waggener Hall, University of Texas, Austin. Texas*
BOSTON	Wilfred S. Lake, *Northeastern University, Boston 15, Massachusetts*
CENTRAL INDIANA	Miss Lucy Kantz, *1504 N. Penn. #5, Indianapolis 2, Indiana*
CENTRAL NEW JERSEY	William Netschert, Jr., *1 Wayside Ave., Lawrence Township, Rt. 101, Trenton, N. J.*
CHICAGO	Randolph T. Snively, *Illinois Bell Telephone Company, Room 1902, 208 West Washington St., Chicago 6, Illinois*
CINCINNATI	Lee M. Welsh, *Procter & Gamble, Gwynne Bldg., Cincinnati 1, Ohio*
CLEVELAND	Gale R. Ober, Jr., *76 South Main St., Chagrin Falls, Ohio*
COLUMBUS	Mikhail V. Condoide, *188 West 10th Avenue, Columbus 1, Ohio*
CONNECTICUT	David F. Votaw, *Yale University, New Haven, Connecticut*
DENVER	Miss Lucy M. Klein, *1205 Washington St., Apt. 22, Denver 3, Colorado*
DETROIT	Andrew T. Court, *1517 Iroquois, Detroit 11, Michigan*
HARRISBURG STATISTICAL SOCIETY	Harry Hoyle, *1824 North 4th St., Harrisburg, Pennsylvania*
HAWAII	Margaret S. Wright, *P. O. Box 1222, Honolulu, T.H.*
ILLINOIS, UNIVERSITY OF	Donald W. Paden, *Department of Economics, University of Illinois, Urbana, Illinois*
ITHACA	Harry J. Loberg, *College of Engineering, Cornell University Ithaca, New York*
LOS ANGELES	William B. Rice, *Box 3519, Terminal Annex, Los Angeles, California*
MADISON, WISCONSIN	Philip J. Fox, *403 Sterling Hall, University of Wisconsin, Madison, Wisconsin*
NEW YORK	Robert E. Lewis, *Federal Reserve Bank of New York, 33 Liberty St., New York 7, New York*
NORTH CAROLINA	George Nicholson, *Department of Mathematical Statistics, University of North Carolina*
OKLAHOMA CITY	Robert S. McCauley, *1710 Exchange, Oklahoma City, Oklahoma*
PACIFIC NORTHWEST	Kermit O. Hanson, *14034 24th Avenue, N.E., Seattle, Washington*
PHILADELPHIA	Kenneth M. Snader, *Federal Reserve Bank of Philadelphia, Philadelphia 1, Pennsylvania*
SACRAMENTO STATISTICAL ASSOCIATION	Marie Vida Ryan, *3551 12th Ave., Sacramento, California*
SAN FRANCISCO	Kenneth C. Christensen, *245 Market Street, San Francisco 6, California*
ST. LOUIS	Kurt Pohlen, *Catholic Hospital Association, 1402 South Grand Boulevard, St. Louis 4, Missouri*
WASHINGTON STATISTICAL SOCIETY	Samuel Weiss, *Bureau of Labor Statistics, Department of Labor, Room 2119. Washington 25, D. C.*

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44 SEPTEMBER 1949 NUMBER 247

ARTICLES

The Monte Carlo Method . . . NICHOLAS METROPOLIS AND S. ULAM 335

Applications of Some Significance Tests for the Median Which Are Valid Under Very General Conditions . . . . . JOHN E. WALSH 342

A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis . . . . . GUY H. ORCUTT AND DONALD COCHRANE 356

Control of a General Census by Means of an Area Sampling Method . . . . . GABRIEL CHEVRY 373

A Procedure for Objective Respondent Selection Within the Household . . . . . LESLIE KISH 380

Beneficiary Statistics Under the Old-Age and Survivors Insurance Program and Some Possible Demographic Studies Based on These Data . . . . . ROBERT J. MYERS 388

By-Product Data and Forecasting in Unemployment Insurance . . . . . NATHAN MORRISON 397

Statistical Requirements for Economic Mobilization. . . . . RALPH J. WATKINS 406

The War Production Board's Statistical Reporting Experience, V and VI . . . . . DAVID NOVICK AND GEORGE A. STEINER 413

BOOK REVIEWS

CROXTON, FREDERICK E. AND COWDEN, DUDLEY J., *Practical Business Statistics, Second Edition* . . . . . ALFRED CAHEN 444

DAHLBERG, GUNNAR, *Mathematical Methods for Population Genetics* . . . . . HOWARD LEVENE 447

EMMENS, C. W., *Principles of Biological Assay* . . LILA F. KNUDSEN 448

GREENSHIELDS, BRUCE D., SCHAPIRO, DONALD, AND ERICKSEN, ELROY L., *Traffic Performance at Urban Street Intersections*. HENRY K. EVANS 451

JEFFREYS, HAROLD, *Theory of Probability, Second Edition* . . . . . HERBERT ROBBINS 453

KENDALL, MAURICE G., *Rank Correlation* . . . . E. J. G. PITMAN 454

PANSE, V. G., *Report on the Scheme for the Improvement of Agricultural Statistics* . . . . . S. LEE CRUMP 455

RICKER, WILLIAM E., *Methods of Estimating Vital Statistics of Fish Populations* . . . . . CHARLES M. MOTTLEY 456

WILKS, S. S., *Elementary Statistical Analysis* . . . . T. A. BANCROFT 458

Letters About Books 460

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 247 SEPTEMBER 1949 *Volume 44*

THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM
Los Alamos Laboratory

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

ALREADY in the nineteenth century a sharp distinction began to appear between two different mathematical methods of treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of *sets of particles*. In pure mathematics an intensive study of the properties of sets of points was the subject of a new field. This is the so-called theory of sets, the basic theory of integration, and the twentieth century development of the theory of probabilities prepared the formal apparatus for the use of such models in theoretical physics, i.e., description of properties of aggregates of points rather than of individual points and their coordinates.

Soon after the development of the calculus, the mathematical apparatus of partial differential equations was used for dealing with the problems of the physics of the continuum. Hydrodynamics is the most widely known field formulated in this fashion. A little later came the treatment of the problems of heat conduction and still later the field theories, like the electromagnetic theory of Maxwell. All this is very well known. It is of course important to remember that the study of the

physics of the continuum was paralleled through "kinetic theories." These consist in approximating the continuum by very large, but finite, numbers of interacting particles.

I

When a physical problem involves an intermediate situation, i.e., a system with a moderate number of parts, neither of the two approaches is very practical. The methods of analytical mechanics do not even give a qualitative survey of the behavior of a system of three mutually attractive bodies. Obviously the statistical-mechanical approach would also be unrealistic.

An analogous situation exists in problems of combinatorial analysis and of the theory of probabilities. To calculate the probability of a successful outcome of a game of solitaire (we understand here only such games where skill plays no role) is a completely intractable task. On the other hand, the laws of large numbers and the asymptotic theorems of the theory of probabilities will not throw much light even on qualitative questions concerning such probabilities. Obviously the practical procedure is to produce a large number of examples of any given game and then to examine the relative proportion of successes. The "solitaire" is meant here merely as an illustration for the whole class of combinatorial problems occurring in both pure mathematics and the applied sciences. We can see at once that the estimate will never be confined within given limits with certainty, but only—if the number of trials is great—with great probability. Even to establish this much we must have recourse to the laws of large numbers and other results of the theory of probabilities.

Another case illustrating this situation is as follows: Consider the problem of evaluating the volume of a region in, say, a twenty-dimensional space. The region is defined by a set of inequalities

$$f_1(x_1, x_2 \cdots x_{20}) < 0; f_2(x_1, x_2 \cdots x_{20}) < 0; \cdots f_{20}(x_1, x_2 \cdots x_{20}) < 0.$$

This means that we consider all points $(x_1, x_2, x_3, \cdots x_{20})$ satisfying the given inequalities. Suppose further that we know that the region is located in the unit cube and we know that its volume is not vanishingly small in general. The multiple integrals will be hardly evaluable. The procedure based on the definition of a volume or the definition of an integral, i.e., the subdivision of the whole unit cube, for example, each coordinate x_1 into ten parts, leads to an examination of 10^{20} lattice points in the unit cube. It is obviously impossible to count all of them. Here again the more sensible approach would be to take, say 10^4 points

at random from this ensemble and examine those only; i.e., we should count how many of the selected points satisfy all the given inequalities. It follows from simple application of ergodic theorems that the estimate should be, *with great probability*, valid within a few per cent.

As another illustration, certain problems in the study of cosmic rays are of the following form. An incoming particle with great energy entering the atmosphere starts a whole chain of nuclear events. New particles are produced from the target nuclei, these in turn produce new reactions. This cascade process continues with more and more particles created until the available individual energies become too small to produce further nuclear events. The particles in question are protons, neutrons, electrons, gamma rays and mesons. The probability of producing a given particle with a given energy in any given collision is dependent on the energy of the incoming particle. A further complication is that there is a probability distribution for the direction of motions. Mathematically, this complicated process is an illustration of a so-called Markoff chain. The mathematical tool for the study of such chains is matrix theory. It is obvious that in order to obtain a mathematical analysis, one would have to multiply a large number of $(n \times n)$ matrices, where n is quite great.

Here again one might try to perform a finite number of "experiments" and obtain a class or sample of possible genealogies. These experiments will of course be performed not with any physical apparatus, but theoretically. If we assume that the probability of each possible event is given, we can then play a great number of games of chance, with chances corresponding to the assumed probability distributions. In this fashion one can study empirically the asymptotic properties of powers of matrices with positive coefficients, interpreted as transition probabilities.

II

Finally let us consider more generally the group of problems which gave rise to the development of the method to which this article is devoted. Imagine that we have a medium in which a nuclear particle is introduced, capable of producing other nuclear particles with a distribution of energy and direction of motion. Assume for simplicity that all particles are of the same nature. Their procreative powers depend, however, on their position in the medium and on their energy. The problem of the behavior of such a system is formulated by a set of integro-differential equations. Such equations are known in the kinetic theory of gases as the Boltzmann equations. In the theory of probabilities one

has somewhat similar situations described by the Fokker-Planck equations. A very simplified version of such a problem would lead to the equation:

$$\frac{\partial u(x, y, z)}{\partial t} = a(x, y, z)\Delta u + b(x, y, z)u(x, y, z) \qquad (1)$$

where $u(x, y, z)$ represents the density of the particles at the point (x, y, z). The Laplacian term, $a\Delta u$ on the right hand side corresponds to the diffusion of the particles, and bu to the particle procreation, or multiplication. [In reality, the equation describing the physical situation stated above is much more complicated. It involves more independent variables, inasmuch as one is interested in the density $w(x, y, z; v_x, v_y, v_z)$ of particles in phase space, v being the velocity vector.] The classical methods for dealing with these equations are extremely laborious and incomplete in the sense that solutions in "closed form" are unobtainable. The idea of using a statistical approach at which we hinted in the preceding examples is sometimes referred to as the Monte Carlo method.

The mathematical description is the study of a flow which consists of a mixture of deterministic and stochastic processes.[1] It requires its own laws of large numbers and asymptotic theorems, the study of which has only begun. The computational procedure looks in practice as follows: we imagine that we have an ensemble of particles each represented by a set of numbers. These numbers specify the time, components of position and velocity vectors, also an index identifying the nature of the particle. With each of these sets of numbers, random processes are initiated which lead to the determination of a new set of values. There exists indeed a set of probability distributions for the new values of the parameters after a specified time interval Δt. Imagine that we draw at random and *independently*, values from a prepared collection possessing such distributions. Here a distinction must be made between those parameters which we believe vary independently of each other, and those values which are strictly determined by the values of other parameters. To illustrate this point: assume for instance that in the fission process the direction of the emitted neutron is independent of its velocity. Or again, the direction of a neutron in a homogeneous medium does not influence the distance between its origin and the site of its first collision. On the other hand, having "drawn" from appropri-

---

[1] von Neumann, J., and Ulam, S., *Bulletin A.M.S.*, Abstract 51-9-165 (1945).

ate distributions the velocity of a new-born particle and the distance to its first collision, the time elapsed in travel is completely determined and has to be calculated accordingly. By considering a large number of particles with their corresponding sets of parameters we obtain in this fashion another collection of particles and a new class of sets of values of their parameters. The hope is, of course, that in this manner we obtain a good sample of the distributions at the time $t+\Delta t$. This procedure is repeated as many times as required for the duration of the real process or else, in problems where we believe a stationary distribution exists, until our "experimental" distributions do not show significant changes from one step to the next.

The essential feature of the process is that we avoid dealing with multiple integrations or multiplications of the probability matrices, but instead sample single chains of events. We obtain a sample of the set of all such possible chains, and on it we can make a statistical study of both the genealogical properties and various distributions at a given time.

III

We want now to point out that modern computing machines are extremely well suited to perform the procedures described. In practice, the set of values of parameters characterizing a particle is represented, for example, by a set of numbers punched on a card. We have at the outset a large number of particles (or cards) with parameters reflecting given initial distributions. The step in time consists in the production of a new such set of cards. The original set is processed one by one by a computing machine somewhat as follows: The machine has been set up in advance with a particular sequence of prescribed operations. These divide roughly into two classes: (1) production of "random" values with their frequency distribution equal to those which govern the change of each parameter, (2) calculation of the values of those parameters which are deterministic, i.e., obtained algebraically from the others. It may seem strange that the machine can simulate the production of a series of random numbers, but this is indeed possible. In fact, it suffices to produce a sequence of numbers between 0 and 1 which have a uniform distribution in this interval but otherwise are uncorrelated, i.e., pairs will have uniform distribution in the unit square, triplets uniformly distributed in the unit cube, etc., as far as practically feasible. This can be achieved with errors as small as desired or practical. What is more, it is not necessary to store a collection of such numbers in the machine itself, but paradoxically enough the machine can

be made to produce numbers simulating the above properties by iterating a well-defined arithmetical operation.

Once a uniformly distributed random set is available, sets with a prescribed probability distribution $f(x)$ can be obtained from it by first drawing from a uniform uncorrelated distribution, and then using, instead of the number x which was drawn, another value $y=g(x)$ where $g(x)$ was computed in advance so that the values y possess the distribution $f(y)$.

Regarding the sequence of operations on a machine, more can be and has been done. The choice of the *kind* of step to be performed by the machine can be made to depend on the values of certain parameters just obtained. In this fashion even dependent probabilistic processes can be performed. Quite apart from mechanized computations, let us point out one feature of the method which makes it advantageous with, say, stepwise integration of differential equations. In order to find a particular solution, the usual method consists in iterating an algebraical step, which involves in the nth stage values obtained from the $(n-1)$th step. The procedure is thus serial, and in general one does not shorten the time required for a solution of the problem by the use of more than one computer. On the other hand, the statistical methods can be applied by many computers working in parallel and independently. Several such calculations have already been performed for problems of types discussed above.[2]

IV

Let us indicate now how other equations could be dealt with in a similar manner. The first, purely mathematical, step is to transform the given equation into an equivalent one, possessing the form of a diffusion equation with possible multiplication of the particles involved. For example as suggested by Fermi, the time-independent Schrödinger equation

$$\Delta\psi(x, y, z) = (E - V)\psi(x, y, z)$$

could be studied as follows. Re-introduce time dependence by considering

$$u(x, y, z, t) = \psi(x, y, z)e^{-Et}$$

u will obey the equation

$$\frac{\partial u}{\partial t} = \Delta u - Vu.$$

[2] Among others, problems of diffusion of neutrons, gamma rays, etc. To cite an example involving the study of matrices, there is a recent paper by Goldberger, *Phys. Rev.* 74, 1269 (1948), on the interaction of high energy neutrons with heavy nuclei.

This last equation can be interpreted however as describing the behavior of a system of particles each of which performs a random walk, i.e., diffuses isotropically and at the same time is subject to multiplication, which is determined by the value of the point function V. If the solution of the latter equation corresponds to a spatial mode multiplying exponentially in time, the examination of the spatial part will give the desired $\psi(x, y, z)$—corresponding to the lowest "eigenvalue" E.

The mathematical theory behind our computational method may be briefly sketched as follows: As mentioned above and indicated by the examples, the process is a combination of stochastic and deterministic flows.[1] In more technical terms, it consists of repeated applications of matrices—like in Markoff chains—and completely specified transformations, e.g., the transformation of phase space as given by the Hamilton differential equations.

One interesting feature of the method is that it allows one to obtain the values of certain given operators on functions obeying a differential equation, without the point-by-point knowledge of the functions which are solutions of the equation. Thus we can get directly the values of the first few moments of a distribution, or the first few coefficients in the expansion of a solution into, for example, a Fourier series without the necessity of first "obtaining" the function itself. "Symbolically" if one is interested in the value of $U(f)$ where U is a functional like the above, and f satisfies a certain operator equation $\psi(f)=0$, we can in many cases obtain an idea of the value of $U(f)$ directly, without "knowing" f at each point.

The asymptotic theorems so far established provide the analogues of the laws of large numbers, such as the generalizations of the weak and strong theorems of Bernoulli, Cantelli-Borel.[3] The more precise information corresponding to that given in the Laplace-Liapounoff theory of additive processes has not yet been obtained for our more general case. In particular it seems very difficult to estimate in a precise fashion the probability of the error due to the finiteness of the sample. This estimate would be of great practical importance, since it alone would allow us to suit the size of the sample to the desired accuracy.

The "space" in which our process takes place is the collection of all possible chains of events, or infinite branching graphs.[4] The general properties of such a phase space have been considered but much work remains to be done on the specific properties of such spaces, each corresponding to a given physical problem.

---

[3] Everett, C. J. and Ulam, S., U.S.A.E.C., Los Alamos reports LADC-533 and LADC-534. Declassified, 1948.

[4] Everett, C. J. and Ulam, S., *Proc. Nat. Acad. Sciences*, 34,403 (1948).

APPLICATIONS OF SOME SIGNIFICANCE TESTS FOR THE MEDIAN WHICH ARE VALID UNDER VERY GENERAL CONDITIONS*

JOHN E. WALSH
The RAND Corporation

In two other papers ([1] and [2]) order statistics were used to derive some tests for the population median which have significance levels which are either exact or bounded under some very general conditions. These order statistic tests were found to be very efficient for small samples from a normal population; also they can be applied with very little computation. This paper contains applications of these tests to several well known statistical problems. Also a graphical method of applying the tests of [1] is outlined.

INTRODUCTION

THE SIGNIFICANCE tests for the population median derived in [1] are valid if the n observations on which a test is based are independent and are drawn from n populations satisfying the conditions

(A)
1) Each population is continuous (i.e. its cumulative distribution is continuous).
2) Each population is symmetrical.
3) The median of each population has the same value ϕ.

These tests compare ϕ with a given constant value ϕ_0.

The tests for comparing ϕ with ϕ_0 derived in [2] are based on the assumption of a sample from a normal population. The significance level of these tests, however, is bounded near the value for normality if the n independent observations are from populations necessarily satisfying only conditions (A).

An important feature of conditions (A) is that no two of the observations are necessarily drawn from the same population. Thus the order statistic tests can be applied to a wide variety of situations.

The motivation for introducing tests with bounded significance levels in addition to the tests of [1] was to obtain a wider variety of suitable significance levels for small values of n without greatly weakening the generality of application.

* The results presented in this paper were obtained in the course of research conducted under the sponsorship of the Office of Naval Research. This research was performed while the author was at Princeton University.

TABLE 1

SOME ONE-SIDED AND SYMMETRICAL SIGNIFICANCE TESTS FOR $n \leq 15$

n	Significance Level of Tests		Tests *Symmetrical:* Accept $\phi \neq \phi_0$ if either		Approx. Efficiency for normality
	One sided	Symmetrical	*One-sided:* Accept $\phi < \phi_0$ if	*One-sided:* Accept $\phi > \phi_0$ if	
4	6.2%	12.5%	$x_1 < \phi_0$	$x_1 > \phi_0$	95%
5	6.2%	12.5%	$\frac{1}{2}(x_1+x_5) < \phi_0$	$\frac{1}{2}(x_1+x_5) > \phi_0$	98%
	3.1%	6.2%	$x_1 < \phi_0$	$x_1 > \phi_0$	96%
6	4.7%	9.4%	$\max\left[x_1, \frac{1}{2}(x_2+x_6)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_5)\right] > \phi_0$	97%
	3.1%	6.2%	$\frac{1}{2}(x_1+x_6) < \phi_0$	$\frac{1}{2}(x_1+x_5) > \phi_0$	98%
	1.6%	3.1%	$x_1 < \phi_0$	$x_1 > \phi_0$	95%
7	5.5%	10.9%	$\max\left[x_1, \frac{1}{2}(x_1+x_7)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_6)\right] > \phi_0$	95%
	2.3%	4.7%	$\max\left[x_2, \frac{1}{2}(x_2+x_7)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_5)\right] > \phi_0$	98%
	1.6%	3.1%	$\frac{1}{2}(x_1+x_7) < \phi_0$	$\frac{1}{2}(x_1+x_5) > \phi_0$	98%
	0.8%	1.6%	$x_1 < \phi_0$	$x_1 > \phi_0$	95%
8	4.3%	8.6%	$\max\left[x_1, \frac{1}{2}(x_1+x_8)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_8)\right] > \phi_0$	94.5%
	2.7%	5.5%	$\max\left[x_2, \frac{1}{2}(x_2+x_8)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_6)\right] > \phi_0$	96%
	1.2%	2.3%	$\max\left[x_3, \frac{1}{2}(x_3+x_8)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_5)\right] > \phi_0$	98%
	0.8%	1.6%	$\frac{1}{2}(x_1+x_8) < \phi_0$	$\frac{1}{2}(x_1+x_5) > \phi_0$	98%
	0.4%	0.8%	$x_1 < \phi_0$	$x_1 > \phi_0$	95%
9	5.1%	10.2%	$\max\left[x_1, \frac{1}{2}(x_1+x_9)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_8)\right] > \phi_0$	91%
	2.2%	4.3%	$\max\left[x_2, \frac{1}{2}(x_2+x_9)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_7)\right] > \phi_0$	96%
	1.0%	2.0%	$\max\left[x_3, \frac{1}{2}(x_3+x_9)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_6)\right] > \phi_0$	95.5%
	0.6%	1.2%	$\max\left[x_4, \frac{1}{2}(x_1+x_9)\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_5)\right] > \phi_0$	99%
	0.4%	0.8%	$\frac{1}{2}(x_1+x_9) < \phi_0$	$\frac{1}{2}(x_1+x_5) > \phi_0$	98%
10	5.6%	11.1%	$\max\left[x_1, \frac{1}{2}(x_1+x_{10})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_9)\right] > \phi_0$	87.5%
	2.5%	5.1%	$\max\left[x_2, \frac{1}{2}(x_2+x_{10})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_8)\right] > \phi_0$	93%
	1.1%	2.1%	$\max\left[x_3, \frac{1}{2}(x_3+x_{10})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_7)\right] > \phi_0$	96.5%
	0.5%	1.0%	$\max\left[x_4, \frac{1}{2}(x_4+x_{10})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_6)\right] > \phi_0$	96.5%
11	4.8%	9.7%	$\max\left[x_1, \frac{1}{2}(x_1+x_{11})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_{10})\right] > \phi_0$	
	2.8%	5.6%	$\max\left[x_2, \frac{1}{2}(x_2+x_{11})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_9)\right] > \phi_0$	89%
	1.1%	2.1%	$\max\left[\frac{1}{2}(x_1+x_{11}), \frac{1}{2}(x_2+x_{10})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)\right] > \phi_0$	
	0.5%	1.1%	$\max\left[x_3, \frac{1}{2}(x_1+x_{11})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_8)\right] > \phi_0$	97%
12	4.7%	9.4%	$\max\left[\frac{1}{2}(x_1+x_{12}), \frac{1}{2}(x_2+x_{11})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)\right] > \phi_0$	
	2.4%	4.8%	$\max\left[x_3, \frac{1}{2}(x_3+x_{12})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_9)\right] > \phi_0$	
	1.0%	2.0%	$\max\left[x_4, \frac{1}{2}(x_4+x_{12})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_7)\right] > \phi_0$	93.5%
	0.5%	1.1%	$\max\left[\frac{1}{2}(x_1+x_{12}), \frac{1}{2}(x_2+x_{10})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_9), \frac{1}{2}(x_2+x_8)\right] > \phi_0$	
13	4.7%	9.4%	$\max\left[\frac{1}{2}(x_1+x_{13}), \frac{1}{2}(x_2+x_{12})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{12}), \frac{1}{2}(x_2+x_{11})\right] > \phi_0$	
	2.3%	4.7%	$\max\left[\frac{1}{2}(x_2+x_{13}), \frac{1}{2}(x_3+x_{12})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{11}), \frac{1}{2}(x_2+x_{10})\right] > \phi_0$	
	1.0%	2.0%	$\max\left[\frac{1}{2}(x_3+x_{13}), \frac{1}{2}(x_4+x_{12})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)\right] > \phi_0$	
	0.5%	1.0%	$\max\left[x_{13}, \frac{1}{2}(x_1+x_{13})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_7)\right] > \phi_0$	94.5%
14	4.7%	9.4%	$\max\left[\frac{1}{2}(x_1+x_{14}), \frac{1}{2}(x_2+x_{13})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{11}), \frac{1}{2}(x_2+x_{10})\right] > \phi_0$	
	2.3%	4.7%	$\max\left[\frac{1}{2}(x_2+x_{14}), \frac{1}{2}(x_3+x_{13})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)\right] > \phi_0$	
	1.0%	2.0%	$\max\left[x_{13}, \frac{1}{2}(x_3+x_{14})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_9)\right] > \phi_0$	90.5%
	0.5%	1.0%	$\max\left[\frac{1}{2}(x_1+x_{14}), \frac{1}{2}(x_{13}+x_{14})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_9), \frac{1}{2}(x_4+x_8)\right] > \phi_0$	
15	4.7%	9.4%	$\max\left[\frac{1}{2}(x_1+x_{15}), \frac{1}{2}(x_2+x_{14})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{13}), \frac{1}{2}(x_2+x_{11})\right] > \phi_0$	
	2.3%	4.7%	$\max\left[\frac{1}{2}(x_2+x_{15}), \frac{1}{2}(x_3+x_{14})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{11}), \frac{1}{2}(x_2+x_{10})\right] > \phi_0$	
	1.0%	2.0%	$\max\left[\frac{1}{2}(x_3+x_{15}), \frac{1}{2}(x_{13}+x_{14})\right] < \phi_0$	$\min\left[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)\right] > \phi_0$	
	0.5%	1.0%	$\max\left[x_{11}, \frac{1}{2}(x_1+x_{15})\right] < \phi_0$	$\min\left[x_1, \frac{1}{2}(x_1+x_9)\right] > \phi_0$	92%

The purpose of this paper is to use these two types of tests to obtain generalized solutions for the cases of quality control, slippage, and the sign test. Also direct applications of the tests are considered.

Tables 1 and 2 contain a list of some practically important one-sided and symmetrical tests of the two types considered ($x_1, \cdots, x_n$ represent the values of the n observations arranged in increasing order of magnitude). The tests of Tables 1 and 2 were chosen so that the significance levels are approximately 5%, 2.5%, 1%, 0.5% for one-sided tests and 10%, 5%, 2%, 1% for symmetrical tests; also so that the amount of computation required for the application of a test is small.

To clarify the use of the tests of Table 1, consider the following example: Let $n=10$ and the values of the 10 observations be

$$0.7,\ -1.1,\ -0.2,\ -1.2,\ 0.1,\ 3.4,\ 3.7,\ 1.3,\ 1.8,\ 2.0.$$

The hypothesis to be tested is that the common median of the continuous symmetrical populations from which these independent observations were drawn has the value zero. Then $\phi_0=0$ and

$$\begin{aligned} x_1 &= -1.2 & x_6 &= 1.3 \\ x_2 &= -1.1 & x_7 &= 1.8 \\ x_3 &= -0.2 & x_8 &= 2.0 \\ x_4 &= 0.1 & x_9 &= 3.4 \\ x_5 &= 0.7 & x_{10} &= 3.7 \end{aligned}$$

Apply the symmetrical test of Table 1 at the 5.1% significance level to these observations. Then

$$\max\left[x_7, \tfrac{1}{2}(x_5 + x_{10})\right] = \max\ (1.8, 2.2) = 2.2 > 0$$
$$\min\left[x_4, \tfrac{1}{2}(x_1 + x_6)\right] = \min\ (0.1, 0.1) = 0.1 > 0.$$

Hence ϕ is significantly different from zero at the 5.1% significance level.

As an example of the application of the tests of Table 2 let $n=6$ and the observations be

$$-0.01,\ 0.21,\ -0.72,\ 0.00,\ -0.05,\ -1.81.$$

The hypothesis to be tested is that the common median of the continuous symmetrical populations from which these six independent observations were drawn has the value 0.1. Then $\phi_0=0.1$ and

$$\begin{aligned} x_1 &= -1.81 & x_4 &= -0.01 \\ x_2 &= -0.72 & x_5 &= 0.00 \\ x_3 &= -0.05 & x_6 &= 0.21. \end{aligned}$$

TABLE 2

SOME ONE-SIDED AND SYMMETRICAL TESTS WITH BOUNDED SIGNIFICANCE LEVELS

Tests n	*SYMMETRICAL:* Accept $\phi \neq \phi_0$ if either		Significance Level for Normality		Significance Level Bounds for Conditions (A)				Approx. Efficiency for normality
	ONE-SIDED: Accept $\phi < \phi_0$ if	*ONE-SIDED:* Accept $\phi > \phi_0$ if	One-sided	Symmetrical	One-sided Tests		Symmetrical Tests		
					Upper	Lower	Upper	Lower	
4	$1.055x_4 - .055x_1 < \phi_0$	$1.055x_1 - .055x_4 > \phi_0$	5%	10%	6.2%		12.5%		96%
5	$.63x_5 + .37x_4 < \phi_0$ $1.02x_5 - .02x_1 < \phi_0$	$.63x_1 + .37x_2 > \phi_0$ $1.02x_1 - .02x_5 > \phi_0$	5% 2.5%	10% 5%	6.2% 3.1%	3.1%	12.5% 6.2%	6.2%	99% 97%
6	$.63x_6 + .37x_5 < \phi_0$ $1.06x_6 - .06x_1 < \phi_0$	$.63x_1 + .37x_2 > \phi_0$ $1.06x_1 - .06x_6 > \phi_0$	2.5% 1%	5% 2%	3.1% 1.6%	1.6%	6.2% 3.1%	3.1%	98.5% 98%
7	$.785x_7 + .215x_6 < \phi_0$ $1.05x_7 - .05x_1 < \phi_0$	$.785x_1 + .215x_2 > \phi_0$ $1.05x_1 - .05x_7 > \phi_0$	1% 0.5%	2% 1%	1.6% 0.8%	0.8%	3.1% 1.6%	1.6%	97% 96%
8	$\max[x_1, (.5x_8 + .28x_6 + .22x_7)] < \phi_0$ $.785x_8 + .215x_7 < \phi_0$	$\min[x_8, (.5x_1 + .28x_8 + .22x_2)] > \phi_0$ $.785x_1 + .215x_2 > \phi_0$	approx. 1% 0.5%	approx. 2% 1%	1.2% 0.8%	0.8% 0.4%	2.3% 1.6%	1.6% 0.8%	98% 97%
9	$\max[x_8, (.5x_8 + .28x_7 + .22x_6)] < \phi_0$	$\min[x_8, (.5x_1 + .28x_8 + .22x_2)] > \phi_0$	approx. 0.5%	approx. 1%	0.6%	0.4%	1.2%	0.8%	98.5%

Apply the symmetrical test of Table 2 which has a 5% significance level if the n observations are a sample from a normal population. Then

$$.63x_6 + .37x_5 = 0.132 + 0.000 = .132 > 0.1$$
$$.63x_1 + .37x_2 = -1.146 - 0.266 = -1.412 < 0.1.$$

Hence the median is not significantly different from 0.1 on the basis of this test. If the six observations are a sample from a normal population, the significance level of the test is 5%. If only conditions (A) are necessarily satisfied, however, the significance level of the test is bounded between 6.2% and 3.1%.

As pointed out in [1], if a symmetrical population has a mean, the mean equals the median. Thus the order statistic tests are tests of the mean if the populations considered have a mean and conditions (A) are satisfied.

The efficiencies listed for the tests in Tables 1 and 2 refer to the power efficiency of the order statistic test considered for the case in which the n observations are a sample from a normal population. The definition of the power efficiency of a test is given in section 3 of [1]. Essentially determination of the power efficiency of a test consists in finding the sample size (not necessarily integral) of the most powerful test of the specified hypothesis (in this case the t-test) which has approximately the same power function as the given test; this sample size divided by the sample size for the given test is called the power efficiency of that test. An intuitive explanation of the meaning of power efficiency is given in [3]. Roughly speaking, the power efficiency of a test is the percentage of the total available information per observation which is being utilized by that test.

Section 2 contains an outline of a graphical method of applying certain of the tests derived in [1]. The remaining sections contain methods of applying the tests listed in Tables 1 and 2 to several well known statistical problems. For simplicity the tests of Tables 1 and 2 will be referred to as the tests of section 1. If tests for $n > 15$ or near different significance levels than those approximated by the tests of Tables 1 and 2 are required, such tests can usually be obtained from the results given in [1].

GRAPHICAL APPLICATION OF TESTS

Let us consider a graphical method applying the tests of Table 1. Since $\max(x, y) < \phi_0$ has the interpretation that both x and y are less than ϕ_0 while $\min(x, y) > \phi_0$ means that both x and y are greater than ϕ_0, it is only necessary to develop a method of deciding how expressions

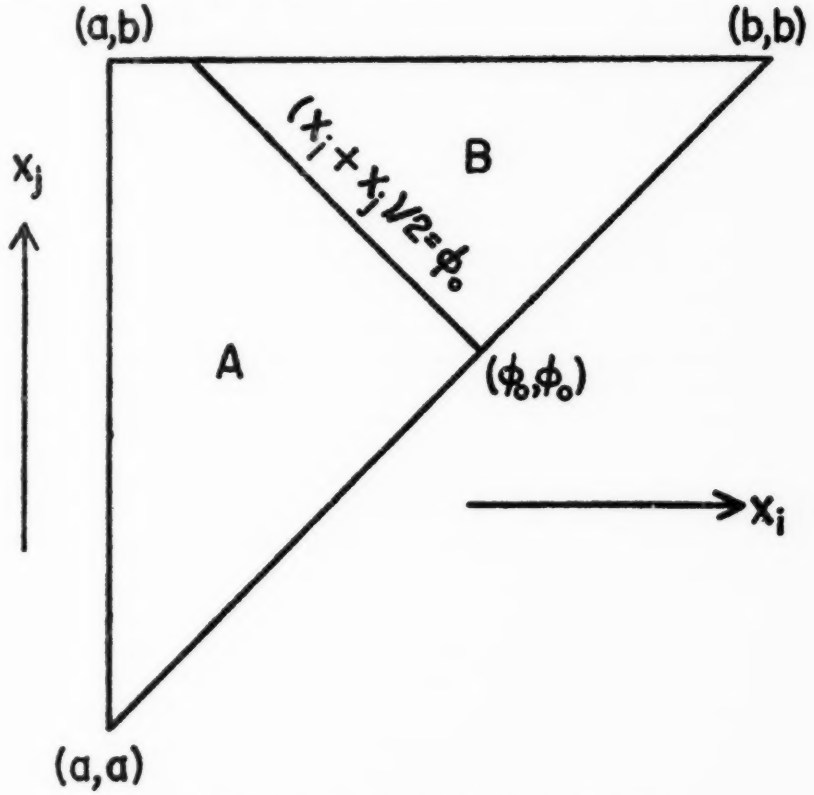


FIG. 1. SCHEMATIC DIAGRAM OF REGIONS A AND B

of the forms x_i and $\frac{1}{2}(x_i+x_j)$ compare with ϕ_0. Direct comparison accomplishes this for the case of x_i. Thus it is sufficient to determine a graphical method of deciding whether $\frac{1}{2}(x_i+x_j)>\phi_0$ or $\frac{1}{2}(x_i+x_j)<\phi_0$.

For nearly all situations to which the tests of Table 1 would be applied, the n observations can be considered to have practical upper and lower limits, say a and b. Using these limits a graphical method of deciding when $\frac{1}{2}(x_i+x_j)<\phi_0$, $(i<j)$, consists in constructing the region A of the x_i, x_j-plane defined by

$$\tfrac{1}{2}(x_i + x_j) < \phi_0, \qquad x_i < x_j, \qquad a \leqq x_i, x_j \leqq b.$$

If the point (x_i, x_j) falls in this region, $\frac{1}{2}(x_i+x_j)<\phi_0$. Similarly $\frac{1}{2}(x_i+x_j)>\phi_0$ if (x_i, x_j) falls into the region B defined by

$$\tfrac{1}{2}(x_i + x_j) > \phi_0, \qquad x_i < x_j, \qquad a \leqq x_i, x_j \leqq b.$$

Fig. 1 contains a schematic diagram of the regions A and B. These regions are particularly easy to construct because the line $\frac{1}{2}(x_i+x_j)=\phi_0$ is perpendicular to the line $x_i=x_j$ at the point (ϕ_0, ϕ_0).

As the regions do not depend on i or j, a single graph will suffice for all the one-sided or symmetrical tests of Table 1. (It is assumed that the bounds a and b are the same for each test.)

DIRECT APPLICATIONS

One important application of the tests of section 1 consists in using these tests as substitutes for the corresponding t-tests. The order statistic tests are more easily applied, valid under more general conditions, and approximately as efficient as the corresponding t-tests.

A second application occurs in cases where it is reasonably certain that the observations are from populations satisfying conditions (A) but there is no reason to suppose that the observations have the same precision. As an example, consider the examination of a gravimetric method of determining the amount of calcium oxide in given samples whose CaO content is known (see [4]). The results for 10 given samples are as follows:

CaO Present	CaO Found By Method	Ratio
Mg.	Mg.	
4.0	3.7	.925
8.0	7.8	.975
12.5	12.1	.968
16.0	15.6	.975
20.0	19.8	.990
25.0	24.5	.980
31.0	31.1	1.003
36.0	35.5	.986
40.0	39.4	.985
40.0	39.5	.988

If it can be assumed that the method used to determine the amount of CaO is symmetrical, the above ratios satisfy conditions (A) with mean (median) equal to unity when the null hypothesis that the average amount of CaO found by the given method equals the true amount holds. Since the amount of CaO present varies from 4.0 mg. to 40.0 mg., the populations from which the ratios are considered drawn may have variances which differ noticeably. Thus application of the t-test to these 10 ratios is a questionable procedure. As conditions (A) are satisfied, however, the tests of section 1 are directly applicable. Apply the symmetrical test of Table 1 at the 1.0% significance level to these 10 ratios. Then $\phi_0=1$ and

$$\max\ [x_9,\ \tfrac{1}{2}(x_6 + x_{10})] = \max\ (0.990,\ 0.994) = 0.994 < 1.000$$

$$\min\ [x_2,\ \tfrac{1}{2}(x_1 + x_5)] = \min\ (0.968,\ 0.950) = 0.950 < 1.000.$$

Thus the method examined yields an average value which is significantly different from the true value at the 1.0% significance level.

USE OF TRANSFORMATIONS

Consider a situation where the n populations from which the n observations were drawn are continuous, have the same median, but are not symmetrical. (A sample from any continuous non-symmetrical population satisfies these conditions.) In practical cases it is sometimes known that replacing each observation value x by the value $g(x)$, where $g(y)$ is a continuous strictly monotonically increasing function of y, will result in a set of observations from approximately symmetrical populations. If ϕ is the common population median for the original observations, $g(\phi)$ will be the population median for the transformed observations. Thus the transformed observations are from populations approximately satisfying conditions (A) with median $g(\phi)$.

Now $g(x_i)$ will be the i^{th} largest of the values of the transformed observations if x_i is the i^{th} largest of the original observations. Also $g(\phi) > g(\phi_0)$ if and only if $\phi > \phi_0$. Similarly for $g(\phi) < g(\phi_0)$ and $g(\phi) \neq g(\phi_0)$. Thus the tests of section 1 are easily modified to obtain tests of $\phi < \phi_0$, $\phi > \phi_0$, and $\phi \neq \phi_0$. The procedure followed is to replace x_i by $g(x_i)$ and ϕ_0 by $g(\phi_0)$ in the body of Tables 1 and 2 but leave everything else as is. For example, for $n = 9$ the one-sided test

Accept $\phi < \phi_0$ *if* $\max[x_8,\ \frac{1}{2}(x_7 + x_9)] < \phi_0$.

is modified to the test.

Accept $\phi < \phi_0$ *if* $\max\ \{g(x_8),\ \frac{1}{2}[g(x_7) + g(x_9)]\} < g(\phi_0)$.

As another example let $n = 8$. Then the one-sided test

Accept $\phi > \phi_0$ *if* $\min\ [x_2,\ (.5x_1 + .28x_3 + .22x_2)] > \phi_0$.

is modified to the test

Accept $\phi > \phi_0$ if $\min\ \{g(x_2),\ [.5g(x_1) + .28g(x_3) + .22g(x_2)]\} > g(\phi_0)$.

The choice of the function $g(y)$ will usually depend on past experience with the type of situation being investigated. For example, in some cases replacing each observation value by the log of that value has been found to yield observations from approximately symmetrical populations.

Since only symmetry is required, there may exist many suitable transformations which have not been used in the past because they do not yield observations from approximately normal populations.

It should be emphasized that tests obtained by transformations are not necessarily also tests of the means of the original populations.

QUALITY CONTROL APPLICATIONS

The tests of section 1 are readily adapted to control chart use. In addition to being valid under extremely general conditions, very efficient for normality, and easily applied, the resulting control chart tests for the median have the valuable property of being independent of the dispersion control chart tests in the sense that the construction and application of the median control charts do not depend on any dispersion values.

Two quality control situations are considered: Control with respect to a given standard value of the median; control with no standard given.

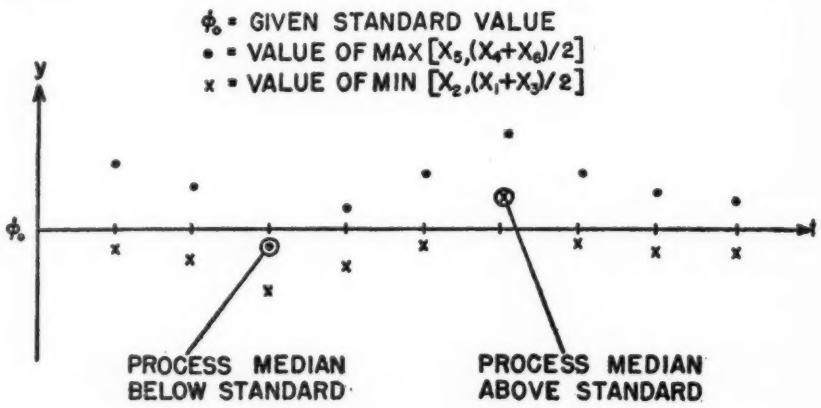


FIG. 2. CONTROL CHART OF MEDIAN FOR SETS OF OBSERVATIONS OF SIZE 6.

As pointed out in section 1, if a symmetrical population has a mean, the mean equals the median. Thus if each observation is drawn from a population satisfying the additional condition that the mean exists, the following quality control tests for the median are also tests for the mean.

A. *Control with respect to given standard.* In this case a standard value ϕ_0 of the median ϕ is given. The method used in constructing and applying control charts based on section 1 tests is demonstrated by the following example: Consider construction of a control chart based on sets of observations of size 6. The time t is plotted on the horizontal axis of the chart while a function y of the six observations is plotted on the vertical axis. The central line on the chart is $y=\phi_0$. The function $y=\max\ [x_5,\ \frac{1}{2}(x_4+x_6)]$ is plotted on the chart with dots while the function $y=\min\ [x_2,\ \frac{1}{2}(x_1+x_3)]$ is plotted with crosses. A dot falling below

the central line indicates that the true value of the median is below the standard value; a cross above the central line indicates that the true value of the median is above the standard value. Fig. 2 furnishes an example of how this control chart might look in application. If conditions (A) hold and the true value of the median equals the given standard value, the probability of a dot falling below the central line equals 4.7%. Similarly for a cross falling above the line. The probability of either a dot falling below the central line or a cross appearing above it is 9.4%. Thus, if the observations are from populations satisfying conditions (A), the control chart exemplified by Fig. 2 represents a graphical method of continually applying a symmetrical test of $\phi \neq \phi_0$ at the 9.4% significance level. If a one-sided test of $\phi < \phi_0$ at the 4.7% significance level is all that is desired, plot only the values of max $[x_5, \frac{1}{2}(x_4+x_6)]$ on the chart. If a one-sided test of $\phi > \phi_0$ at the 4.7% significance level is sought, plot only the values of min $[x_2, \frac{1}{2}(x_1+x_3)]$.

The method of constructing control charts outlined above is directly applicable to all the section 1 tests. As another example, consider sets of size 5. Let a dot be defined as $1.02x_5-.02x_1$ and a cross by $1.02x_1-.02x_5$. Then Fig. 2 represents a continual application of a symmetrical test of $\phi \neq \phi_0$ with significance level 5% for normality and upper bound 6.2% for conditions (A). Plotting $1.02x_1-.02x_5$ alone furnishes a one-sided test of $\phi > \phi_0$ with significance level 2.5% for normality and upper bound 3.1% for conditions (A), etc.

Control charts based on unequal size sets of observations can also be readily obtained by use of the tests of section 1. The method used to obtain control charts for these cases consists in giving a separate definition of what a dot and cross are to represent for each set size. The significance level can differ according to size or be approximately the same for all sizes. As an example of the procedure used, consider a control

TABLE 3

DEFINITION OF DOTS AND CROSSES

Set Size	Definition of Dot	Definition of Cross
5	$1.02x_5-.02x_1$	$1.02x_1-.02x_5$
6	$.63x_6+.37x_5$	$.63x_1+.37x_2$
7	max $[x_6, (x_5+x_7)/2]$	min $[x_2, (x_1+x_3)/2]$
8	max $[x_6, (x_5+x_8)/2]$	min $[x_2, (x_1+x_4)/2]$
9	max $[x_7, (x_6+x_9)/2]$	min $[x_3, (x_1+x_5)/2]$
10	max $[x_7, (x_6+x_{10})/2]$	min $[x_4, (x_1+x_6)/2]$

chart using sets varying from 5 to 10 observations in size. Let the dots and crosses for Fig. 2 be defined by Table 3. Then, if conditions (A) hold, Fig. 2 represents a continual application of a symmetrical test of $\phi \neq \phi_0$ at approximately the 5% significance level. The dots alone furnish a one-sided test of $\phi < \phi_0$ at approximately the 2.5% level while the crosses by themselves represent a one-sided $\phi > \phi_0$ at approximately the 2.5% significance level.

B. *Control—no standard given* (*control chart test*). For this case a standard value of the median is not given. The standard value is replaced by an estimate of the true median value made on the basis of past data taken while the process was in control. Once a suitable estimate of the median is obtained, the determination of control charts for this case becomes identical with that considered in section A if ϕ_0 is replaced by the estimated value of the median. Hence the main problem is to obtain a suitable estimate for the median on the basis of the given past observations.

A very satisfactory estimate can be obtained if the past observations were drawn from populations satisfying conditions (A) and the additional condition that the first three moments of each population are finite. Then the average of all the past observations furnishes an estimate of the true value of the median. This estimate has the favorable properties:

1. The expected value of the estimate equals the true value of the median.
2. The estimate tends to the true median value as the number of past observations increases.

From an application viewpoint, the additional condition that the first three moments of each population are finite is not very restrictive; this condition is satisfied for nearly all populations arising in practice.

Use of the above estimate in place of ϕ_0 allows the control chart methods developed in section A to be utilized.

GENERALIZED SLIPPAGE TEST

The usual slippage situation investigated is the following: A change is made with affects a continuous population in such a way that the shape of the population distribution remains fixed but the population mean may move. This slippage of the mean is tested on the basis of samples drawn from the population both before and after the change.

The purpose of this section is to generalize the above slippage problem and present a solution to the generalized situation. The generalized slippage problem is the following: A change is made which affects m

continuous populations; k of these populations, $(k=0, 1, \cdots, m)$, are affected in such a way that each population has the same shape distribution before and after the change; the remaining $m-k$ populations are symmetrical both before and after the change (but the distribution shapes may change). It is required to test whether the change affected the values of the means of the m populations.

The method used to derive this test consists in obtaining n dummy observations, $(n \geqq 6)$, which satisfy conditions (A) with zero median when the null hypothesis that all the means remained fixed (no slippage) is true. Slippage tests with a wide variety of significance levels can then be obtained by applying tests based on conditions (A) to these dummy observations. In particular, if $n \leqq 15$, the tests of Tables 1 and 2 of section 1 can be used.

The procedure used to obtain this test is the following:

(a) $1 \leqq m \leqq 5$. Choose r such that $6 \leqq rm \leqq 15$. Then draw r samples of size s from each population both before and after the change. Record the order in which these samples were drawn. Form the mean of each sample. Consider the mean of the i^{th} sample drawn from the j^{th} population, $(i=1, \cdots, r; j=1, \cdots, m)$, after the change. Subtract this mean from the mean of the i^{th} sample drawn from that population before the change. Under the conditions of the generalized slippage problem it is easily seen that the resulting rm dummy observations satisfy conditions (A) with zero median if there is no slippage.

(b) $m \geqq 6$. Draw a sample of size s from each population both before and after the change. Form the mean of each sample. Subtract the mean of the sample drawn from the j^{th} population after the change, $(j=1, \cdots, m)$, from the mean of the sample drawn from that population before the change. The resulting m dummy observations satisfy conditions (A) with zero median if the null hypothesis of no slippage is true.

In both cases (a) and (b) a set of independent dummy observations satisfying conditions (A) with zero median are obtained under the null hypothesis of no slippage. The slippage test is obtained by applying the appropriate section 1 test to these dummy observations.

The tests can be generalized by replacing the condition that all samples drawn are of the same size by more general conditions. Modifications of this nature can be made in several obvious ways and will not be discussed here. Also in case (a) it is possible to choose r such that $rm > 15$. The restriction $rm \leqq 15$ was imposed so that the tests of Tables 1 and 2 could be used.

GENERALIZED SIGN TEST

A test which has wide application is the well known sign test (see, e.g. [5]). This test is used to compare two kinds of objects (say X and Y) with respect to a specified characteristic. The comparison is accomplished by pairing the two types of objects. In each pair the value of the characteristic obtained for the Y object is subtracted from the value obtained for the X object. This procedure furnishes n dummy observations $x_1-y_1, \cdots, x_n-y_n$, where n is the number of pairs formed. The conditions under which the observations $(x_1, y_1), \cdots, (x_n, y_n)$ where obtained are such that the dummy observations can be assumed to be independent and to have the property.

$$(1) \qquad Pr\,(x_i - y_i > 0) = Pr\,(x_i - y_i < 0) = \tfrac{1}{2}, \qquad (i = 1, \cdots, n),$$

if the null hypothesis of no difference between X and Y with respect to the specified characteristic is true. The sign test is then applied to test this null hypothesis on the basis of the signs (plus or minus) of the n dummy observations.

Let us examine condition (1) from the viewpoint of approximate verification in practice. The main practical situations (i.e., situations approximately satisfied in practice) from which condition (1) can be deduced occur when there is reason to believe that each observation (x_i, y_i) satisfied one or more of the following conditions when the null hypothesis is true:

(a) The joint cumulative distribution $F(x_i, y_i)$ of x_i and y_i is continuous and such that $F(x_i, y_i) = F(y_i, x_i)$; i.e., x_i and y_i receive identical treatment.

(b) x_i and y_i are independent and are from continuous symmetrical populations with the same median value (the two populations are not necessarily the same).

(c) x_i and y_i have the same population median and a normal bivariate distribution.

In all of cases (a)–(c), the dummy observation x_i-y_i is from a continuous symmetrical population with zero median (see [6]). Thus, from the viewpoint of practical verification, assuming that the n dummy observations satisfy condition (1) is almost equivalent to assuming that the dummy observations satisfy conditions (A) with zero median. This suggests that the tests of section 1 be used for those practical situations where the sign test is ordinarily used. The section 1 tests are preferable to the sign test from the viewpoint of suitable significance levels and efficiency for normality.

It would be very convenient if $x-y$ were from a continuous sym-

metrical population with zero median whenever the observation (x, y) satisfies the condition

(d) x and y come from symmetrical populations with the same median value; also the joint cumulative distribution of x and y is continuous.

Then conditions (b) and (c) could be replaced by the single condition (d) and the problem of deciding when $x-y$ is from a continuous symmetrical population with zero median would be considerably simplified. The following counter-example, however, shows that condition (d) is not sufficient to ensure that $x-y$ is from a continuous symmetrical population with zero median:

Let the joint probability density function of x and y be defined by

$$f(x, y) = \frac{1}{4} + \frac{x}{10}(1 - 3y^2) \quad \text{if} \quad -1 \leqq x, y \leqq 1$$

$$= 0 \qquad\qquad\qquad\qquad \text{otherwise.}$$

Integration shows that the marginal distributions of x and y are both symmetrical with zero median. However $Pr(x-y<0)$ is easily shown to not equal $\frac{1}{2}$.

REFERENCES

[1] John E. Walsh, "Some significance tests for the median which are valid under very general conditions," *Annals of Math. Stat.*, March, 1949.

[2] John E. Walsh, "On the range-midrange test and some tests with bounded significance levels," *Annals of Math. Stat.*, June, 1949.

[3] John E. Walsh, "On the 'information' lost by using a t-test when the population variance is known," *Jour. Amer. Stat. Assoc.*, March, 1949.

[4] Wallace M. Hazel and Warren K. Eglof, "Determination of calcium in magnesite and fused magnesia," *Ind. Eng. Chem. Anal. Ed.*, Vol. 18 (1946), pp. 759–760.

[5] W. J. Dixon and A. M. Mood, "The statistical sign test," *Jour. Amer. Stat. Assoc.*, Vol. 41 (1946), pp. 565–566.

[6] John E. Walsh, "Some significance tests for the median which are valid under very general conditions." Unpublished thesis, Princeton University.

A SAMPLING STUDY OF THE MERITS OF AUTOREGRESSIVE AND REDUCED FORM TRANSFORMATIONS IN REGRESSION ANALYSIS

GUY H. ORCUTT AND DONALD COCHRANE
Department of Applied Economics, Cambridge

This paper is concerned with some aspects of regression analysis when the error terms are autocorrelated and there exists more than one relationship between the variables. In particular, we investigate the merits of autoregressive transformations and the reduced form transformation in dealing with these complications.

An important result is that, unless it is possible to specify something about the intercorrelation of the error terms in a set of relations and to choose approximately the correct autoregressive transformation, a certain amount of scepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observations.

1. INTRODUCTION

THE STATISTICAL estimation of the various parameters which enter into theoretical formulations of economic relationships is one of the main objectives of econometrics and the most common statistical technique used is multivariate regression analysis. The classical method of least squares regression has been shown to give best linear unbiassed estimates of the coefficients when certain well known conditions are fulfilled. If a linear relationship exists between the dependent variable x_{1t} and a set of independent variables $x_2 \cdots x_p$ of the form

$$x_{1t} = b_0 + \sum_{j=2}^{p} b_{1j}x_{jt} + u_t \tag{1}$$

these conditions are satisfied if among other things[1]

(i) the error term is non-autocorrelated, so that the expected value $E(U_t \cdot U_{t-h}) = 0$ for $h \neq 0$;

(ii) each of the determining variables x_{jt} $(j=2, \cdots, p)$ is independent of the error term U_t, i.e., $E(x_{jt}U_t) = 0$ $(j=2, \cdots, p)$.

[1] For a complete statement of the conditions under which least squares give "best unbiassed" estimates, see F. N. David and J. Neyman, "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs*, Vol. II, London 1938.

The formidable complications which have arisen in estimating the structural parameters of economic relationships have their origin, in so far as they are purely statistical in nature, in the fact that these two conditions are not realistic and have to be relaxed in most applications to economic data. The major complications which have arisen may be classified as follows:

(a) the auto-correlated error complication;
(b) the errors in variables complication;
(c) the simultaneous equations complication.

The first of these complications arises when condition (i), that the errors are independently distributed in time, does not hold while condition (ii), that the determining variables are independent of the error term, cannot be maintained when the other two complications are present.

The simultaneous equations approach. The data used in most formulations of economic relationships are obtained from historical time processes and not from conducted experiments, and as a consequence are the results of the solution of a system of simultaneous relations corresponding to the economic processes involved. In order to obtain accurate estimates of the structural parameters of any single equation it may therefore be necessary to take account of the whole system of simultaneous equations in which it occurs. Consider a simple illustration of this problem. The consumption and price of a commodity enter into both a demand and a supply relation so that if we attempt to find the demand relation by considering only the regression of the quantity consumed on the price of the commodity we are ignoring the fact that price is not an independent variable but will depend on the nature of the supply relation. Haavelmo[2] suggested that in such cases the variables should be considered in a joint normal probability distribution which should be studied to clarify the stochastical relationship which the system of equations implies. Such a method assumes that the errors in the equations are non-autocorrelated and normally distributed and that there are no errors of observation in the variables.

It has been shown that for large samples the parameters estimated by the method of maximum likelihood from this joint probability distribution have certain optimal properties. They are asymptoticallyunbiassed estimates and are also efficient statistics.[3] One method of esti-

[2] T. Haavelmo, "The Probability Approach in Econometrics," Supplement to *Econometrica*, Vol. 12, July 1944.

[3] A more complete discussion can be found in T. Koopmans, "Statistical Methods of Measuring Economic Relationships," *Cowles Commission Discussion Papers* Statistics No. 310 (mimeographed copy of lectures delivered at the University of Chicago 1947) and T. Haavelmo, "Methods of Measuring the Marginal Propensity to Consume," *Journal of the American Statistical Association*, Vol. 42, 1947, pp. 105–122.

mating these parameters is to rewrite the system of equations in the reduced form and solve for each of the endogenous variables in terms of the lagged values of the endogenous variables and the exogenous variables which appear in the system. These solutions will be in terms of linear equations and the method of least squares can then be applied by considering each endogenous variable in turn as the dependent variable. The coefficients in this form possess the properties of best unbiassed estimates for large samples. The structural parameters of the original equations can then be derived from these coefficients but it should be mentioned that it may not always be possible to identify the structural parameters of the original relations from the coefficients of the equations estimated in the reduced form. The problem of identification is a very important one for the method discussed and a careful analysis of the system of equations that is being considered should be made before attempting any statistical application.[4] The estimates obtained from the reduced form method are maximum likelihood solutions for an exactly identified system.

Autocorrelated error terms. In an earlier paper[5] we showed that the error terms involved in many current formulations of economic relationships are highly positively autocorrelated. In doing so, we demonstrated that under these circumstances the application of least squares regression to the original data produced very inefficient estimates of the parameters to be measured and suggested that this efficiency could be recovered by applying an autoregressive transformation to the variables which would make the error term approximately random.

Objects of this paper. In this paper we are concerned with the problem of carrying out regression analysis when the error terms are autocorrelated and there exists more than one relationship between the variables. In particular we investigate the merits of autoregressive transformations and the reduced form transformation in dealing with these complications. It is assumed that there are no errors of observation in the variables.

The problems with which we are dealing are essentially deductive in nature and the ideal solution to them is one reached by purely deductive steps from stated premises. However, since it has not been possible

---

[4] See Koopmans, "Statistical Methods of Measuring Economic Relationships," *op. cit.* A very good description of the practical procedure is contained in M. A. Girshick and T. Haavelmo, "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations," *Econometrica*, Vol. 15, 1947, pp. 79–110. Further references may be found in the various articles to which we have referred.

[5] D. Cochrane and G. H. Orcutt, "Application of Least Squares Regression to Relationships containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 32–61.

as yet to obtain such a solution, and considering the problems of some importance, we have resorted to the method of sampling experiments. That is, we embody our assumptions in experimental models, use these models to generate sets of time series and investigate empirically the results of various estimating procedures on the series generated. Up to the present only large sample properties of the parameters derived by the simultaneous estimation of structural relations have been demonstrated. Since economic data rarely comprise series of more than 20 years, these large sample properties would seem to require more careful investigation and the sampling experiment approach provides a convenient and legitimate method of making such an investigation. In addition it might be mentioned that the use of such methods might also provide the answers to many problems which have proved intractable to mathematical statistics and the improvements in calculating equipment are very welcome for these purposes.

2. CONSTRUCTION OF THE EXPERIMENTAL MODELS

In order to reduce the computational burden as much as possible we worked with the simplest types of systems which seemed at all reasonable from the standpoint of applying any conclusions to economic studies. Two models were adopted and are explained as follows. The original series of Model I were generated by a recursive system of equations

$$x_t = a_0 + a_1 y_t + (u_{1t} + u_{2t}) \tag{2}$$

$$y_t = b_0 + b_1 x_{t-1} + u_{3t} \tag{3}$$

where x_t and y_t are the series to be considered and u_{it} $(i = 1, 2, 3)$ are the error terms involved in the two relations. These error terms were generated by the autoregressive equations

$$u_{it} = u_{i,t-1} + \epsilon_{it} \qquad (i = 1, 2, 3) \tag{4}$$

where the ϵ_{it} denote series of random disturbances. The values of the parameters in (2) and (3) were chosen to be

$$\begin{cases} a_0 = b_0 = 0 \\ a_1 = 1.0 \\ b_1 = 0.4 \end{cases} \tag{5}$$

The ϵ_{it} $(i = 1, 2, 3)$ are independently distributed single digit random numbers. They were extracted from *Tables of Random Sampling*

Numbers,[6] ignoring zeros so that they ranged from 1 to 9. Subtracting the number 5 from each we obtained three random series ϵ_{it}, possessing rectangular distributions with ranges of +4 to −4 and an expected value of zero. The three series u_{it} were then generated by applying equations (4) with initial values of zero, so that we had three independent series of first summations of random elements each comprising over 500 terms. Taking x_0 as zero and making use of the properties of the system given by (2) and (3), we then generated long series of x_t and y_t from the three error series u_{it}. The first five items of the series of x_t and y_t were discarded and the remaining long series were each divided into 20 segments of 21 items with 5 items omitted between segments. One of these items was later used for prediction. By this procedure we obtained a sample of 20 pairs of series generated by the same underlying autoregressive structure but involving different samples of random disturbances.

For Model II the original series were generated by the same process as just described for Model I except that instead of being independent the error terms were now highly intercorrelated. This result was achieved by using the same series of u_{1t} and u_{2t} as before but replacing the series u_{3t} by u_{2t} so that the true correlation between the error terms was 0.71. The recursive system therefore became

$$x_t' = a_0 + a_1 y_t' + (u_{1t} + u_{2t}) \tag{6}$$

$$y_t' = b_0 + b_1 x_{t-1}' + u_{2t} \tag{7}$$

where the constants remained the same as given in (5). The same procedure was used to obtain the individual sets of x_t and y_t as explained for Model I.

Choice of parameters. Our choice of the autoregressive properties of the error series u_{it} was based upon the evidence presented in our previous paper[7] and the reasonableness of assuming that error terms are first summations of random elements has been further supported from the results obtained by Stone for a number of demand studies in the United Kingdom.[8] Our choice of the product $a_1 b_1$ was made so that x_t and y_t would have approximately the same autoregressive

[6] M. G. Kendall and B. Babington-Smith, "Tables of Random Sampling Numbers," *Tracts for Computers No. 24*, Cambridge University Press 1939.

[7] D. Cochrane and G. Orcutt, *op. cit.*

[8] Richard Stone, "The Analysis of Market Demand: An Outline of Methods and Results" read before a meeting of the European section of the Econometric Society at The Hague, September 1948, and to be published in *The Review of the International Statistical Institute.*

structures as claimed by Orcutt[9] for the series used in Tinbergen's[10] model of the economic system of the United States. For instance in the case of independent error terms we can see from (2) and (3)that the autoregressive structures of the two series are

$$x_t = x_{t-1} + 0.4(x_{t-1} - x_{t-2}) + \eta_{1t} \tag{8}$$

$$y_t = y_{t-1} + 0.4(y_{t-1} - y_{t-2}) + \eta_{2t} \tag{9}$$

where η_{1t} and η_{2t} are random disturbances defined in terms of $\epsilon_{it}(i = 1, 2, 3)$.

Having made the decision as to the product a_1b_1, only one more significant decision remains to be made about the general structure of the model. This is the correlation between either pair of variables x_t and y_t or y_t and x_{t-1}. For n approaching infinity this may be more clearly seen as follows. We may express our model in the form

$$x_t = ay_t + v_{1t} \tag{10}$$

$$y_t = bx_{t-1} + v_{2t} \tag{11}$$

where x_t and y_t are in terms of deviations from their means and v_{1t}, v_{2t} are random error series. Expressing x_t and y_t in autoregressive forms, we can derive the following relations:

$$\frac{E(x_t^2)}{E(v_{1t}^2)} = \frac{1}{1 - \rho^2}\left(1 + a^2\frac{E(v_{2t}^2)}{E(v_{1t}^2)}\right) = R_1 \tag{12}$$

$$\frac{E(y_t^2)}{E(v_{2t}^2)} = \frac{1}{1 - \rho^2}\left(\frac{\rho^2}{a^2}\,\frac{E(v_{1t}^2)}{E(v_{2t}^2)} + 1\right) = R_2 \tag{13}$$

where

$$ab = \rho, \quad R_1 = \frac{1}{1 - r_{x_t y_t}^2} \quad \text{and} \quad R_2 = \frac{1}{1 - r_{y_t x_{t-1}}^2}.$$

It can be readily seen that, having decided the correlation between say x_t and y_t, we have R_1 and since the term

$$a^2\frac{E(v_{2t}^2)}{E(v_{1t}^2)}$$

[9] G. H. Orcutt, "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States 1919–32," *Journal of the Royal Statistical Society*, Vol. X, Series B, 1948, pp. 1–53.

[10] J. Tinbergen, "Statistical Testing of Business-Cycle Theories Vol. II; Business Cycles in the United States of America 1919–32," League of Nations, Geneva, 1939.

appears in both (12) and (13), then R_2 is a function of ρ and R_1 which are both known and is automatically determined. All we have left to decide is the weights to be assigned to the coefficient a and the relative variances of the error terms. We made the relative variances

$$E(v_{2t}^2)/E(v_{1t}^2) = \tfrac{1}{2},$$

so that for $r_{x_t y_t}^2 = 0.44$ we have $a = a_1 = 1$. The resulting form provides an intermediate and reasonable model of a simplified economic system. On the basis of the calculations needed for this study, it is possible to work out the extreme cases in which either of the error terms has zero variance. This is done in the next section.

3. CALCULATIONS INVOLVED AND SOME SPECIAL CASES

The results of the calculations carried out are contained in Tables 1 to 5. The equations referred to in these tables are:

Model I.

$$\text{I.} \quad x_t = a_0 + a_1 y_t + (u_{1t} + u_{2t})$$

(14) $$\text{II.} \quad y_t = b_0 + b_1 x_{t-1} + u_{3t}$$

$$\text{III.} \quad x_t = \rho_0 + \rho_1 x_{t-1} + (u_{1t} + u_{2t} + a_1 u_{3t}).$$

Model II.

$$\text{IV.} \quad x_t' = a_0 + a_1 y_t' + (u_{1t} + u_{2t})$$

(15) $$\text{V.} \quad y_t' = b_0 + b_1 x_{t-1} + u_{2t}$$

$$\text{VI.} \quad x_t' = \rho_0 + \rho_1 x_{t-1}' + (u_{1t} + (1 + a_1) u_{2t}).$$

Both these systems are exactly identified. Equations I and IV were calculated by the direct application of least squares regression. Equations II and V are already in the reduced form so that the use of least squares is the appropriate procedure. The reduced forms of equations I and IV are equations III and VI respectively, so that the reduced form estimates of a_1 and a_0 are given by

(16) $$a_1' = \rho_1 / b_1$$

(17) $$a_0' = \rho_0 - b_0 a_1' = \bar{x}_t - a_1' \bar{y}_t$$

where $\bar{x}_t$ and $\bar{y}_t$ are the means of the two series x_t and y_t. The calculations relating to these estimates are given by equations IA and IVA in Model I and Model II respectively. For each of the equations we have made first and second difference transformations and the regression parameters have been estimated in the three forms. The original

relations and the autoregressive transformations are denoted by the letters O, F.D. and S.D. in the tables.

Special cases. At the end of section 2 we pointed out that it is possible to derive the cases where the variance of either one of the error terms is equal to zero. In both these cases the least squares estimates and the reduced form estimates lead to identical results. Rewriting the simple system of (10) and (11)

$$x_t = ay_t + v_{1t} \tag{18}$$

$$y_t = bx_{t-1} + v_{2t} \tag{19}$$

where v_{1t}, v_{2t} are random elements so that the reduced form of x_t is

$$x_t = \rho x_{t-1} + v_{1t} + av_{2t} \qquad (\rho = ab) \tag{20}$$

we can say that when (19) is an exact relation then the single equation least squares estimates and the reduced form estimates of a are the same and proportional to the estimate of the autoregressive coefficient obtained in (20). When (18) is an exact relation then the single equation least squares and the reduced form estimates of a are both exact. These equalities may be more clearly seen as follows:

First assumption $E(v_{2t}^2) = 0$

so that our system becomes

$$x_t = ay_t + v_{1t} \tag{21}$$

$$y_t = bx_{t-1} \tag{22}$$

$$x_t = \rho x_{t-1} + v_{1t} \tag{23}$$

and the least squares estimate of b is exact.

The single equation least squares estimate of a is given by

$$\hat{a} = \frac{\sum x_t y_t}{\sum y_t^2} = \frac{1}{b} \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2} = \frac{\hat{\rho}}{b} \tag{24}$$

where $\hat{\rho}$ is the least squares estimate of ρ. The reduced form estimate of a is therefore

$$\tilde{a} = \frac{\hat{\rho}}{b} = \hat{a}. \tag{25}$$

Second assumption $E(v_{1t}^2) = 0$

so that our system is now

$$x_t = ay_t \tag{26}$$

$$y_t = bx_{t-1} + v_{2t} \tag{27}$$

$$x_t = \rho x_{t-1} + av_{2t} \tag{28}$$

and the single equation least squares estimate of a is exact. Now (27) and (28) are identical except for a scalar multiplier a, therefore the reduced form estimates of a are $\hat{p}/\hat{b}$ and will also be exact.

4. GENERAL RESULTS

In the ensuing discussion we shall be content to point out the general and more important features of the calculations and if the reader desires further information it may be obtained from the tables which are presented in detail.

Structural parameters. It has been proved by Mann and Wald[11] that for large samples a linear stochastic difference equation may be treated as a classical regression problem in which the lagged values of the series appear as independent variables. However, the adequacy of the ordinary least squares regression has not been demonstrated for small samples, particularly of the size usually considered by economists, and in fact Koopmans[12] has mentioned that for a sample of 3 items are bias will be present in the least squares estimates of the parameters of a single lag autoregressive equation. Orcutt[13] has pointed out that this bias is probably due partly to the necessity of using the sample means of the time series instead of the true means and partly to the skewness of the distribution of sample estimates even when the true means are used. His empirical evidence shows that this bias is very substantial for series having only a weak central tendency. If we look at equations III and VI in Table 1 we find, for the first difference transformation, examples of single lag autoregressive equations. In these cases the means of the estimated regression coefficients are given, and it can be seen that the biasses are 2.3 and 2.8 times the standard error of the means of twenty estimates. This indicates that even for low values of the autoregressive coefficient the bias is still rather large. When we estimate the coefficients assuming a true mean of zero we find from equations III and VI in Table 2 that the bias is considerably reduced but is still far from negligible.

The question naturally arises as to whether a similar sort of small sample bias is to be expected in single equation least squares and reduced form estimates of the parameters of a system of recursive equations. In the previous section we showed that when either one of the equations was exact the estimates of both methods were the same. They were exact in one case and possessed the same bias and variance as the

[11] H. B. Mann and H. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, Vol. 11, 1943, pp. 173–220.

[12] T. Koopmans, "Serial Correlation and Quadratic Forms in Normal Variables," *Annals of Mathematical Statistics*, Vol. 13, 1942, pp. 14–33.

[13] G. H. Orcutt, *op. cit.*

TABLE 1

REGRESSION PARAMETERS

Equation	Autoregressive transformation		Constant term (True value zero)			Regression Coefficient						Correlation coefficient			Estimated mean variance of error
									Variance						
	Single equation	Reduced form	Mean	Standard error of mean	Variance	True value	Mean	Standard error of mean	Using mean	Using true value	Mean of estimated variance	Mean	Standard error of mean	Variance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Model I															
I	O		−25.75	13.62	3712.	1.0	1.27	0.11	0.223	0.283	0.038	0.84	0.05	0.057	20.6
	F.D.		− 0.18	0.16	0.54	1.0	0.97	0.05	0.053	0.052	0.082	0.62	0.03	0.014	14.1
	S.D.		− 0.04	0.08	0.11	1.0	0.61	0.07	0.094	0.239	0.129	0.40	0.03	0.021	28.2
IA		O	35.11	116.91	298030.	1.0	2.10	0.48	4.502	5.494	—	—	—	—	83.4
		F.D.	− 0.63	1.05	30.11	1.0	0.62	0.15	0.432	0.551	—	—	—	—	18.1
		S.D.	0.18	0.69	10.15	1.0	−1.82	0.66	8.605	16.104	—	—	—	—	198.3
II	O		−10.57	9.55	1824.	0.4	0.59	0.05	0.058	0.091	0.007	0.82	0.05	0.045	11.4
	F.D.		− 0.25	0.15	0.43	0.4	0.36	0.03	0.014	0.015	0.017	0.54	0.03	0.018	6.8
	S.D.		− 0.07	0.03	0.02	0.4	0.20	0.03	0.021	0.061	0.023	0.29	0.04	0.037	12.3
III	O		−18.36	4.71	444.	0.4	0.90	0.02	0.008	0.255	0.013	0.88	0.02	0.006	22.8
	F.D.		− 0.57	0.24	1.19	0.4	0.26	0.06	0.062	0.079	0.050	0.25	0.05	0.057	21.4
	S.D.		0.01	0.09	0.15	0.4	−0.27	0.05	0.050	0.490	0.053	−0.27	0.05	0.049	30.3
Model II															
IV	O		−14.28	7.77	1206.	1.0	1.48	0.06	0.079	0.301	0.016	0.94	0.01	0.004	13.3
	F.D.		− 0.07	0.10	0.183	1.0	1.51	0.03	0.020	0.276	0.054	0.83	0.02	0.005	11.2
	S.D.		− 0.03	0.07	0.097	1.0	1.53	0.03	0.072	0.345	0.140	0.70	0.02	0.006	28.2
IVA		O	− 8.23	9.59	1839.	1.0	1.52	0.08	0.114	0.378	—	—	—	—	14.4
		F.D.	− 0.54	0.26	1.334	1.0	0.52	0.18	0.668	0.863	—	—	—	—	24.2
		S.D.	3.03	2.04	83.038	1.0	5.23	7.85	1226.5	1883.1	—	—	—	—	17470.1
V	O		− 8.07	5.94	760.	0.4	0.59	0.03	0.018	0.055	0.004	0.89	0.02	0.007	10.0
	F.D.		− 0.13	0.12	0.279	0.4	0.33	0.02	0.011	0.014	0.011	0.59	0.03	0.020	7.4
	S.D.		− 0.06	0.05	0.042	0.4	0.09	0.02	0.006	0.099	0.012	0.19	0.03	0.020	11.7
IV	O		−28.13	5.34	570.	0.4	0.86	0.02	0.009	0.224	0.014	0.86	0.02	0.005	36.2
	F.D.		− 0.33	0.23	1.082	0.4	0.23	0.06	0.062	0.086	0.050	0.23	0.06	0.061	34.9
	S.D.		− 0.15	0.11	0.249	0.4	−0.34	0.04	0.036	0.583	0.051	−0.34	0.04	0.034	47.7

coefficient of a single lag autoregressive equation of the type just considered in the other. It is therefore of interest to examine first our Model I which corresponds to an intermediate case from the two extremes and second our Model II which adds the complication due to intercorrelated error terms and accordingly a further bias to the single equation estimates by the direct correlation between the independent variable and the error series.

First look at the single equation estimates of the regression coefficients given in Table 1, by the sets of equations I and IV. The estimates based on the original series are badly biassed and have large variances in both methods. As expected the bias is greater in Model II where the error terms are intercorrelated. The mean of the variances of the regression coefficients estimated from each equation separately (see column 11) does not reflect the true position and is only a fraction of what it should be. When we make a first difference transformation the estimates of the regression coefficient in Model I possess very little bias while the mean of the estimated variances of the regression coefficients also appears to be reasonable. However, in Model II there is still a large bias due to the correlation of the independent variable and the error term of equation IV.

Turning to the reduced form estimates given by equations IA and IVA we see that they are badly biassed for both the original series and first differences of both models. In the case of the original series the

TABLE 2
REGRESSION PARAMETERS CALCULATED BY ASSUMING TRUE MEAN OF ZERO
(First Difference Transformation)

Equation	Regression Coefficient					Correlation Coefficient		
	True Value	Mean	Standard error of mean	Variance		Mean	Standard error of mean	Variance
				Using mean	Using true value			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model I								
I	1.0	1.00	0.05	0.041	0.039	0.66	0.03	0.014
IA	1.0	0.79	0.12	0.290	0.319	—	—	—
II	0.4	0.39	0.03	0.015	0.014	0.58	0.03	0.017
III	0.4	0.33	0.05	0.059	0.061	0.33	0.05	0.056
Model II								
IV	1.0	1.50	0.03	0.021	0.271	0.84	0.01	0.004
IVA	1.0	0.66	0.13	0.352	0.446	—	—	—
V	0.4	0.35	0.02	0.009	0.011	0.61	0.03	0.017
IV	0.4	0.27	0.05	0.056	0.070	0.27	0.05	0.055

biasses are in the same direction as the single equation biasses but are much larger, in the first difference transformation the bias is downwards in both models and is due to the short series bias previously considered.

So far we have considered the coefficients obtained when we estimated the means in each transformation. It is therefore of interest to see the bias in the estimates of the first difference transformation when we make use of the fact that the true mean of the series is zero. This is equivalent to assuming that there is no trend in the original relationships. The results for the first difference transformation of both models are given in Table 2. In the case of Model I the estimates obtained by the single equation least squares regression are not biassed but the same bias as previously obtained is present in the case of equation IV for Model II. The reduced form estimates are still biassed in both models, although they show an improvement over the estimates obtained when the means are estimated.

Our calculations may be used to see whether the means of the reduced form estimates are significantly biassed from the true values in the first difference transformations. When we calculate the coefficient using estimated means we find from Table 1 that the values 0.62 from equation IA and 0.52 from equation IVA are both significantly different from the true value of the regression coefficient at the 5 per cent. level,[14] using the standard error of the mean calculated from the variance around the estimated mean. In fact they are significantly different from the true value of the coefficient at the 2 per cent level. When we assume that the true means of the series are zero only the mean of the reduced form estimates for Model II is significantly biassed. From Table 2 it can be seen that the value of 0.79 for equation IA is not significant at the 5 per cent. level but the value of 0.66 for equation IVA is significant at the 2 per cent level.

Consider now the efficiency of the least squares and reduced form estimates. In Table 1 we find that the variances of the reduced form estimates compare very unfavourably in all cases with the variances of the single equation estimates. This may be illustrated by the ratios of the reduced form variances to the single equation variances. Even when we include the effect of the bias on the estimates by calculating the variance around the true value of the regression coefficient, the ratios still remain very high for Model I and although they fall slightly in Model II they are still greater than unity. For the cases

[14] Using the t tables for 20 degree of freedom. We are using both tails of the distribution but a case can be made for using only one tail. This would halve the levels of significance.

	Ratios of variances regression estimates	
	Using estimated mean	Using true value
Model I		
Original series	20:1	19:1
First differences	8:1	11:1
Model II		
Original series	1½:1	1½:1
First differences	30:1	3:1

where we assume a knowledge of the true mean, Table 2 shows that the ratio of the variances of the reduced form estimates to the variances of the single equation estimates around the estimated means and the true value are 7:1 and 8:1 respectively for Model I and 17:1 and 1½:1 for Model II. These are still very high ratios and are surprising results.

In Model I, where the error terms are random and independent, the single equation estimates would be the maximum likelihood solutions for large samples and normally distributed error terms[15] and we therefore expected the single equation approach to produce better estimates than those obtained by the reduced form method. However, where the error terms are random and intercorrelated the reduced form estimates are the maximum likelihood solutions for large samples if the correlation between the unlagged error terms is unknown. This is the case of Model II and we should therefore have expected the results of Model I to be reversed, but the discussion of the last few paragraphs has shown that these expectations have been far from realised.

Autocorrelation of residuals. The autocorrelations of the true error terms and the estimated residuals are presented in Table 3. The measure of autocorrelation used is the ratio of the mean square successive difference to the variance. This statistic is usually denoted by δ^2/s^2 and for a random series it is symmetrically distributed around a mean of $2n/n-1$ where n is the number of items in the series.[16] In a previous paper[17] we showed that highly positively autocorrelated error terms become strongly biassed towards randomness as the number of para-

[15] R. Bentzel and H. Wold, "On Statistical Demand Analysis from the viewpoint of Simultaneous Equations," *Skandinavisk Aktuarietidskrift*, Vol. 29, 1946, pp. 95–114.

[16] The probability distribution of this statistic has been tabulated; see B. S. Hart and J. von Neumann, "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Varian," *Annals of Mathematical Statistics*, vol. 13, pp. 207–214.

[17] D. Cochrane and G. Orcutt, *op. cit.*

meters in the estimation relationship increases. This result is again illustrated in Table 3.

Further it should be noted that not only are the residuals biassed when the single equation least squares method of estimation is used, but this bias appears to be approximately the same as can be seen from (column 7) Table 3, when the reduced form method of estimation is used.

TABLE 3
AUTOCORRELATION OF RESIDUALS

Equation	Autocorrelation transformation		Values of δ^2/s^2						
			Actual error series				Estimated residuals		
	Single equation (1)	Reduced form (2)	Values for infinite series (4)	Mean (5)	Standard error of mean (5)	Variance (6)	Mean (7)	Standard error of mean (8)	Variance (9)
Model I									
I	O		0.0	0.61	0.09	0.14	1.00	0.10	0.22
	F.D.		2.0	2.16	0.10	0.20	2.21	0.10	0.21
	S.D.		3.0	—	—	—	—	—	—
IA		O	0.0	0.61	0.09	0.14	1.07	0.10	0.19
		F.D.	2.0	2.16	0.10	0.20	1.93	0.09	0.18
		S.D.	3.0	—	—	—	—	—	—
II	O		0.0	0.43	0.06	0.08	0.97	0.11	0.22
	F.D.		2.0	2.05	0.09	0.17	2.00	0.08	0.14
	S.D.		3.0	—	—	—	—	—	—
III	O		0.0	—	—	—	1.49	0.10	0.21
	F.D.		2.0	—	—	—	1.87	0.05	0.04
	S.D.		3.0	—	—	—	—	—	—
Model II									
IV	O		0.0	0.61	0.09	0.14	1.10	0.13	0.34
	F.D.		2.0	2.16	0.10	0.20	2.50	0.10	0.20
	S.D.		3.0	—	—	—	—	—	—
IVA		O	0.0	0.61	0.09	0.14	1.04	0.11	0.24
		F.D.	2.0	2.16	0.10	0.20	2.07	0.11	0.24
		S.D.	3.0	—	—	—	—	—	—
V	O		0.0	0.51	0.06	0.08	1.17	0.10	0.18
	F.D.		2.0	2.17	0.09	0.17	2.01	0.08	0.11
	S.D.		3.0	—	—	—	—	—	—
IV	O		0.0	—	—	—	1.47	0.11	0.22
	F.D.		2.0	—	—	—	1.92	0.06	0.06
	S.D.		3.0	—	—	—	—	—	—

There is also an additional bias in the residuals caused by the application of biassed estimates of the regression coefficients. This may be seen in the residuals of the first difference transformation in equa-

tion IV, Table 3 (column 7). It can be easily shown that the application of biassed coefficients in this equation will produce negatively autocorrelated error terms and this is illustrated in the result obtained. A final interesting feature of the autocorrelation of the residuals is to be seen in equations III and VI. In the first difference transformation we are estimating the coefficient of a single lag autoregressive equation with random disturbance but it is noticeable that because of the downward bias in the estimate of the coefficient the residuals are not completely randomized.

Prediction. In each equation we forecasted the dependent variable

TABLE 4

ERRORS OF FORECAST

Equation	Autoregressive transformation		Actual errors of forecast			Estimated mean variance of	
	Single equation	Reduced form	Mean	Standard errors of mean	Variance	Error	Individual forecast
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model I							
I	O		2.45	1.09	23.8	20.6	24.0
	F.D.		−0.03	0.65	8.6	14.1	16.2
	S.D.		−1.52	1.10	24.3	28.2	31.3
IA		O	6.19	4.57	417.1	83.4	—
		F.D.	−1.03	0.76	11.5	18.1	—
		S.D.	−5.86	3.43	235.7	198.3	—
II	O		−1.27	1.04	21.6	11.4	13.8
	F.D.		−1.06	0.67	9.0	6.8	7.6
	S.D.		−1.12	0.68	9.2	12.3	13.9
III	O		0.63	1.35	36.6	22.8	26.7
	F.D.		−1.18	1.04	21.9	21.4	23.8
	S.D.		−2.30	1.18	27.8	30.3	34.0
Model II							
IV	O		1.92	1.09	25.6	13.3	15.4
	F.D.		−0.66	0.59	7.0	11.2	12.4
	S.D.		−1.22	1.19	28.2	28.2	31.0
IVA		O	1.65	1.14	26.0	14.4	—
		F.D.	0.38	0.91	16.4	24.2	—
		S.D.	40.56	34.39	23654.	17470.	—
V	O		0.38	0.74	10.9	10.0	11.6
	F.D.		0.02	0.53	5.6	7.4	8.3
	S.D.		−0.77	0.68	9.2	11.7	13.1
IV	O		2.10	1.07	23.0	36.2	41.4
	F.D.		0.31	1.03	21.0	34.9	38.9
	S.D.		−1.96	1.24	30.7	47.7	53.9

for the next period from a knowledge of the independent variable and using the regression coefficients calculated from the series up to that period. The forecasts were then compared with the actual value of the dependent variable obtained from our constructed series as explained in section 2, and the variances of error are shown in Table 4. The forecasts based directly upon the single equation estimates of the parameters have a smaller variance of error than those based on the direct use of the reduced form estimates of the parameters in the set of structural equations IA and IVA. This is true for both models and each autoregressive transformation. In both models the smallest variance of error of forecast is given when using single equation estimates in the first difference transformation.

We also calculated the variance of forecast for the two main equations of both models in first difference transformation using the knowledge that the true means of the series were zero. Table 5 shows that these results do not change the general impression and the single equation least squares estimates still provide the smallest variances of the errors of forecast.

TABLE 5

ERRORS OF FORECAST ASSUMING TRUE MEAN OF ZERO

(First Difference Transformation)

Equation	Actual errors of forecast		
	Mean	Standard error of mean	Variance
Model I			
I	−0.10	0.58	6.66
IA	0.36	0.66	8.64
Model II			
IV	0.41	0.54	5.82
IVA	−0.69	0.90	16.15

5. CONCLUSIONS

Care must be taken when drawing general conclusions from sampling experiments of the type considered in this paper but the results appear to us to be of a striking and significant nature. They indicate that unless it is possible to specify with some degree of accuracy the intercorrelation between the error terms of a set of relations and unless it is possible to choose approximately the correct autoregressive transformation so as to randomize the error terms, then a certain amount of scepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observa-

tions when generated by systems analogous to those examined in this paper. This scepticism will be considerably increased if it is also attempted to make a choice of variables and time lags from the same data

If the error terms are independent or nearly independent between equations, the results obtained justify the use of single equation least squares estimation, in which case the main problem lies in making the correct autoregressive transformation. This would seem to be the situation as regards many demand relationships, particularly those which are in terms of current variables and those relating to agricultural products.

For short run prediction we would appear to be in a somewhat more favourable position since the estimated variance of errors of individual forecasts obtained from single equation regression analysis seem to be in line with the actual errors and as small as could be expected even in the absence of a simultaneous equations complication. However, it must be remembered that such predictions assume that the same system will continue.

CONTROL OF A GENERAL CENSUS BY MEANS OF AN AREA SAMPLING METHOD

GABRIEL CHEVRY
Institut National de la Statistique et des Études Économiques, Paris

The area sampling method was applied to industrial and commercial enterprises in urban areas of France to determine the adequacy of the general census in enumerating them. The investigation showed an under-reporting by the census of enterprises of all sizes.

A GENERAL CENSUS of the population was taken in France on March 10, 1946. Like the previous ones, this census involved in effect several simultaneous inquiries about people, households, dwellings, industrial and commercial enterprises, and agricultural establishments; different questionnaires were used for each of these categories of statistical units.

Questionnaire No. 5 applied to the industrial and commercial enterprises answering the following definition: an enterprise consists of a group of two or more persons working together permanently, in a permanent place, under the management of one or more representatives of the same trade name. According to this definition, a person working alone did not constitute an enterprise. Two partners, or a husband and wife working together, without assistants, constituted an enterprise employing no salaried staff. Each of the various subsidiaries of the same firm constituted a separate enterprise, even if they were located in the same community.

A first reckoning of the questionnaires No. 5 which had been filled through the census procedure revealed some anomalies, which indicated that more than a negligible number of enterprises had not filled their answers to the census questionnaires. In order to have a more precise impression and to be able to appraise quantitatively the accuracy of the census of industrial and commercial enterprises, and management of the Direction des Enquetes economiques of the Institut national de la Statistique decided in October 1946 to resort to its regional offices for an *a posteriori* control of the census process by means of an area sampling method.

PRINCIPLES OF THE METHOD

To reduce the cost of the process, the control was limited to the

urban areas in which the Institut national de la Statistique has regional offices. It could be presumed that the census of March 10, 1946, had been more accurately taken in the rural areas than in the large cities.

The control consisted simply in selecting at random in each of these cities, a certain number of areas in which the sampling was to take place, in listing the enterprises located in the selected areas, and in sending investigators to visit all these areas to make an inventory of all the enterprises located therein and to obtain completed questionnaires No. 5 from those enterprises which had not submitted their answers to the census.

For each city the number of areas selected was considered adequate when, according to the census data, the number of enterprises located in these areas was at least equal to 2.5 per cent of the total number of enterprises located in the whole urban center.

TECHNIQUE OF SAMPLING

The first step consisted in setting the limits of the urban centers where the sampling was to take place. It was required that the sampling not be limited to the city in which the regional office of the Institut national was located, but should extend to the whole urban center in which the city was included, this center consisting of all the small communities or bounding parts of these communities in which the inhabited lots of these divisions were adjoining or joined together by parks, gardens, orchards, yards, workshops or other similar enclosures, even if these houses or enclosures were separated from each other by a ditch, a river or a public garden.

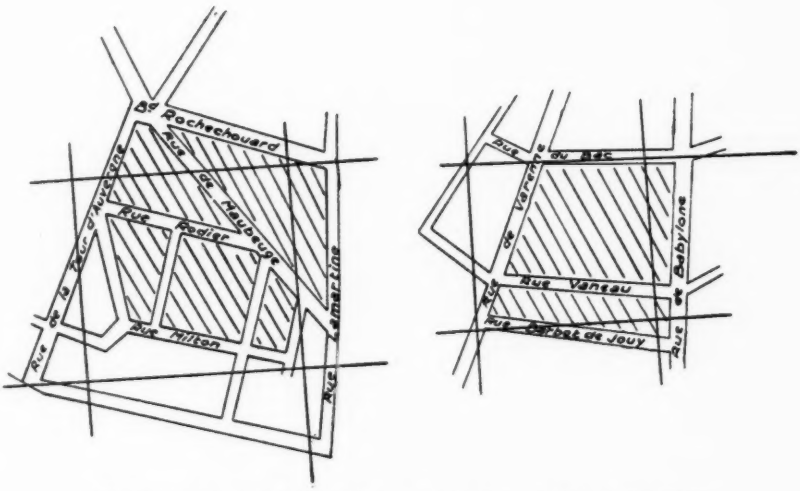
A large scale map of the urban center thus delineated was covered with a regular orthogonal grid forming squares of 250 meters per side for the smaller centers and 500 meters per side for the very large ones. The squares covering areas containing construction were numbered and a certain number of squares were taken at random by means of Tippett tables.

As it was impossible to locate on the terrain the limits of the selected grids, instructions were given to substitute for each of these grids, polygons of approximately equal areas but bounded by actual streets. The chart below shows two examples of this substitution, which was indispensable in order to: a) facilitate the listing of the enterprises located in each of the grids which had filled the questionnaire; and b) point out unambiguously the buildings which were to be visited by the investigators.

All the buildings located in the polygons thus selected and delineated

were visited by the investigators. Each of the investigators was provided with a list of the enterprises which, according to the census, were located in the group of buildings assigned to him. He checked on this list the enterprises which had completed answers to Questionnaire No. 5 for the census authorities, and those which had not been polled by the census.

As this procedure took place in November 1946, several months after the general census, special instructions were given to the investigators,



SUBSTITUTION OF POLYGONS FOR SQUARES. AREAS INCLUDED IN POLYGONS ARE SHOWN AS SHADED REGIONS

so that they could distinguish, among all of the enterprises existing in November 1946 and not having replied to the census questionnaire, those which had been in existence in March 1946 and thus should have been included in the census list, and those which had been created or reopened between March and November. Conversely, it was hardly possible to discover in November the number of enterprises which had been closed since March. This fact seems of no consequence, however, as bankruptcies were very scarce in this period after the war.

RESULTS OF THE SAMPLING

Table 1 shows the gross results of the sampling for each of the selected cities and for the sample as a whole.

In the 18 urban centers where the sampling control was used,

TABLE 1

GROSS RESULTS OF THE SAMPLE

Urban Centers	Total number of enterprises included in the census of March 1946 in these centers	Number of enterprises existing in the sample			
		Included in the census of March 1946		Not included in the census of March 1946	
		Number	As a percentage of the total number of enterprises in the census	Number	As a percentage of the number of enterprises in the sample
(1)	(2)	(3)	(4) = (3)/(2)	(5)	(6) = (5)/[(3)+(5)]
			%		%
Bordeaux	10,556	365	3.4	97	21.0
Clermont	3,185	127	4.0	52	29.1
Dijon	2,366	60	2.5	29	32.6
Lille	7,549	228	3.0	188	45.2
Limoges	2,870	105	3.6	34	24.5
Lyon	14,230	405	2.8	369	47.6
Marseille	14,355	531	3.7	182	25.5
Montpellier	3,281	161	4.9	61	27.5
Nancy	3,272	88	2.7	27	23.5
Nantes	5,005	195	3.9	19	8.9
Orleans	2,748	109	3.9	17	13.5
Paris	110,000	3,115	2.8	529	14.5
Poitiers	905	53	5.9	38	41.8
Reims	3,233	120	3.7	9	7.0
Rennes	2,835	89	3.1	23	20.5
Rouen	5,351	160	3.0	87	35.2
Strasbourg	5,958	128	2.2	60	31.9
Toulouse	4,580	115	2.5	82	41.6
Total	201,359	6,154	3.1	1,903	23.6

201,359 enterprises had been included in the census of March 1946. The selected sample included 6,154 enterprises, i.e. 3.1 per cent of the total.

The investigation made on the spot has shown that 8,057 enterprises existed in this sample. This means that 1,903 enterprises, or an amount equal to 30.9 per cent of the enterprises included in the census and 23.6 per cent of the existing enterprises, had been overlooked in the census of March 1946.

Sampling theory permits one to determine the accuracy of this gross result. It does not seem possible to use unreservedly the formula which, in the case of a cluster sample gives the variance of an estimate of probability, p; (p, in this case, would be the probability of an enterprise being overlooked in the census). This formula supposes in effect

that all the clusters are of the same size, i.e. include the same number of enterprises. It is obvious that, with the method applied in the sample, this condition has not been fullfilled.

To measure the accuracy of the results under these circumstances, it seems more advisable to apply the approximate formula provided by Goldberg.[1] This formula deals with the errors in estimating a ratio x/y, when the ratio is obtained as the quotient of estimates of its numerator and denominator, both computed from at least n elements. This estimate implies a bias whose average relative value is

$$\frac{C_x^2 - \rho C_x C_y}{n}$$

and the relative standard error of sampling is

$$\sqrt{\frac{C_x^2 + C_y^2 - 2\rho C_x C_y}{n}}$$

where

C_x is the coefficient of variation of x, namely $\sigma_x/\bar{x}$,
$\bar{x}$ is the arithmetical mean of the values of x in the universe C_y,
C_y is the coefficient of variation of y namely $\sigma_y/\bar{y}$, and
ρ is the coefficient of correlation between x and y.

In the particular case which we are considering, n is the number of polygons, bounded by streets, which were taken at random; the values of y are the number of the enterprises existing in each of these polygons, and the values of x are the number of enterprises overlooked by the census in the same polygons.

The coefficient of correlation is rather high: 0.91. The average relative value of the bias is 2/1000, i.e. practically negligible. The relative value of the standard error of the sample is 7.6 per cent.[2]

It can therefore be estimated that in 95 cases out of 100 the percentage of enterprises overlooked in the census lies within the limits of $23.6(1-2\times0.076)$ or 20 per cent, and $23.6(1+2\times0.076)$ or 27.2 per cent. Thus we can state that, in the census of March 10, 1946, 20 to 27 per cent of the industrial and commercial enterprises were overlooked.

It has already been pointed out that *a priori* it does not seem sound

[1] Cf: "Méthodes statistiques modernes des Administrations fédérales aux États-Unis" by P. Thionet (Hermann & Cie, Paris, 1946) p. 58, and "Sampling theory when the sampling units are of unequal sizes" by W. G. Cochran (*Journal of the American Statistical Association*, June 1942).

[2] Actually, these calculations could be made for only 14 out of the 18 cities in which the sampling took place. In the other 4 cities, we could not use the results relative to each of the selected polygons. This in no way nullifies the conclusions which can be drawn from the results, however, as the degree of error which was based on the data for 14 towns was greater than that found for the 18 towns.

to apply this result to the entire French territory; the sampling has been limited to the large cities and there is no reason to believe that the census of enterprises has been as incomplete in the rural areas as in the cities. Furthermore the high correlation between x and y might indicate that the converse is true; namely the denser the population of enterprises, the greater the number of omissions.

COMPLEMENTARY RESULTS

Although it may be interesting to know that in the large cities 20 to 27 per cent of the enterprises were overlooked in the general census, this information is nevertheless insufficient. One must ask oneself how and why the census has been incomplete. In particular, the first question that presents itself is the following: Have the enterprises been overlooked in the census because of their nature or their size? One might in fact think that principally the small enterprises have been omitted.

An answer to this question, at least with regard to the size of the enterprises, is furnished by Table 2, which compares the distribution according to the number of their salaried employees, of the 756,235

TABLE 2

DISTRIBUTION, ACCORDING TO THE NUMBER OF THEIR SALARIED EMPLOYEES, OF THE 756,235 ENTERPRISES INCLUDED IN THE CENSUS AND THE 1,623 ENTERPRISES NOT INCLUDED IN THE CENSUS BUT REVEALED BY THE SAMPLE

Number of salaried employees	Enterprises included in the census		Enterprises not included in the census but appearing in the sample	
	Number	Per cent of total	Number	Per cent of total
0	230,354	30	397	25
1	197,252	26	546	33
2 to 5	214,504	29	443	27
6 to 10	48,561	6	100	6
11 to 20	29,188	4	65	4
21 to 50	21,294	3	47	3
51 to 100	8,001	1	15	1
More than 100	7,081	1	10	1
Total	756,235	100	1,623	100

enterprises included in the census of March 1946 covering France's whole territory, with an analogous distribution of the 1,623 enterprises which the sample revealed were omitted from the census.

The two series of percentages show a very satisfactory agreement for all the enterprises having 2 or more salaried employees. For the two

other catgeories (0 and 1 salaried employee) the agreement is less satisfactory. It does not seem necessary, however, to attach much importance to this fact. First of all, the percentages resulting from the sample can be considered as only approximate, in view of the standard deviation of the sample. Moreover, the distribution of enterprises between these two categories is subject to some uncertainty. The enterprises comprising only two persons could be classified either as an enterprise consisting of two employers and consequently no salaried employee, or as one consisting of one employer and one salaried employee, according to the manner in which the informant interpreted the questionnaire. It is, in addition, quite possible that, particularly for tax purposes, an enterprise having 2 employers and no salaried employee had claimed having one employer and one employee, e.g. that the employer's wife had been erroneously recorded as a salaried assistant.

Under these circumstances, it seems wiser to put together the first two categories of Table 2. In this way, one obtains percentages of 56 per cent for the census and 58 per cent for the sample, which may be considered as sufficiently alike.

Thus it seems possible to conclude that, if the census of 1946 has been inadequate, it has been so for the enterprises of all sizes. According to the information gathered by the investigators of the sample, the fact that such a high percentage of enterprises was overlooked in the census was attributable to negligence on the part of the census enumerators.

For these reasons the Institut national de la Statistique has not published the results of the 1946 Census of industrial and commercial enterprises. Since then, others methods have been used to examine this subject. A permanent inventory of these enterprises has been established. It will give a more complete and exact picture of the industrial and commercial structure of France.

A PROCEDURE FOR OBJECTIVE RESPONDENT SELECTION WITHIN THE HOUSEHOLD*

LESLIE KISH
Survey Research Center, University of Michigan

In modern survey methods growing emphasis is placed on the objective selection of the sample. For surveys of the general population, increasing use is made of area sampling to obtain probability samples of households. Heretofore, scant attention has been given to the question of how to make an objective selection among the members of the household.

A procedure for selecting objectively one member of the household is given as used in four surveys of the adult population. Demographic data as found in the sample are compared with outside sources for available factors.

The Problem

TO OBTAIN random samples for surveys,[1] two basic conditions must be satisfied:

1. The sampling method must provide for a known probability of inclusion, other than zero, for every element of the population.
2. The method must be translated into a procedure that can and will be applied in practice.

Area sampling is gaining acceptance as a practical and reliable procedure for obtaining samples of households with known probabilities of selection.[2] In general practice these probabilities are equal either within specified strata or throughout the sample.

Given a sampling procedure—such as area sampling—which uses dwellings as units of sampling: when does the question of selection within households arise? There is no need of selection:

---

* Presented at the 107th Annual Meeting of the American Statistical Association, New York City, December 30, 1947.

[1] For discussion of sources of biases that may arise when personal judgment enters into the selection see: Hauser, Philip M. and Hansen, Morris H. "On Sampling in Market Surveys," *Journal of Marketing*, July 1944, pp. 26–31.

[2] King, A. J. and Jessen, R. J., "The Master Sample of Agriculture," *Journal of the American Statistical Association*, Vol. 40, No. 229, March, 1945, pp. 38–42.

Hansen, M. H. and Hauser, P. M., "Area Sampling—Some Principles of Sample Design," *Public Opinion Quarterly*, Summer, 1945, pp. 183–193.

Houseman, Earl E., "The Sample Design for a National Farm Survey by the Bureau of Agricultural Economics," *Journal of Farm Economics*, Vol. XXIV, No. 1, February, 1947, pp. 241–245.

"A Chapter in Population Sampling" by the Sampling Staff of the Census Bureau, 141 pages, Superintendent of Documents.

A. If the respondent is uniquely determined (as head of household or homemaker); or

B. If the household is the unit of analysis and any adult member can give equally valid information.[3]

However, if the household contains more than one member of the desired population, it may be regarded as a cluster of population elements.

One may decide to include in the sample every member of the population within the household.[4] However, this may be a statistically inefficient procedure in general (depending on the cost-variance relationship of the survey design), unless one of these three conditions holds:

A. Information about all members can be obtained from one of them.

B. There is seldom more than one member of the population in the household. For example: there is an average of 1.2 spending units per household.[5]

C. If the intra-class correlation within the household of the variables measured is of negligible size, or if it is negative.

These conditions are not met generally in surveys of the attitudes of the adult population. Furthermore, multiple interviews in one household may lead to undesirable interview situations. Hence, there is need for a procedure of selection that will translate a sample of the households into a sample of the adult population. There are no great theoretical difficulties, but a practical procedure must meet the demands of efficiency and of applicability. There are several alternatives, and the choice among them depends on a number of factors: the nature and distribution of the population, the objectives and design of the survey, and the available facilities. Under the latter we may distinguish the factors of cost, and of the nature and training of the field force.

The Conditions

It may be useful to describe very briefly the general sampling procedures used by the Survey Research Center.[6] A stratified random

---

[3] This assumption may be unjustified. See: Deming, W. Edwards, "On Errors in Surveys," *American Sociological Review*, August, 1944, p. 361.

[4] See, for example: Watson, Alfred N., "Respondent Pre-Selection within Sample Areas," No. 2 of Technical Series on Statistical Methods in Market Research, page 7, Research Department, Curtis Publishing Company, Philadelphia.

[5] Based on the Consumer Finances Surveys of the Federal Reserve Board, conducted by the Survey Research Center.

[6] For a fuller description see: Goodman, Roe, "Sampling for the 1947 Survey of Consumer Finances," *Journal of the American Statistical Association*, September, 1947, pp. 439–448.

selection of dwellings is obtained by means of area sampling. Small areas, segments, define the clusters of households within counties (or groups of adjacent counties) selected as primary sampling areas. In each of these areas there are trained interviewers who are employed on a part-time basis. Detailed sampling instructions are given to them in order to insure correct understanding and execution of the different procedures required by a variety of types of surveys.

The interviews are of the fixed question, free-answer kind, requiring generally from 30 minutes to one and one-half hours. Conducting the interviews in the home is believed to contribute toward a satisfactory interview situation. The interviewer calls at pre-designated dwellings; the respondent is then selected by a fixed procedure. Return calls are made to find the not-at-homes.

For each of a series of four surveys, conducted in June, August and December, 1946 and April, 1947 respectively, it was desired to obtain samples of about 600 interviews of the adult population of the Continental United States. The samples were distributed in 31 primary sampling areas: in four of the 12 largest metropolitan areas plus 27 scattered counties.

The population of the four surveys was limited to adults living in private households, excluding, because of interviewing difficulties and cost considerations, some segments of the population: armed forces; hospitals; religious, educational and penal institutions; trailer, logging and labor camps; and hotels and large rooming houses.

A procedure was devised for selecting one adult in each sample household. This procedure was favored over some alternatives (such as selecting every other adult found in sample households) for two reasons:

1. It was desired to take no more than one interview in any household, in order to obtain each interview before the respondent had a previous opportunity to discuss the questions. Furthermore, multiple interviews were thought to be statistically inefficient because of the expected correlation of attitudes within the household.
2. An interview in every sample household was desired in order to avoid making futile calls on dwellings without interviews.

With this procedure unbiased estimates may be obtained by giving each respondent as weight the number of adults in the household. Such differential weighting may in general increase the sampling error. However, in the present instance this increase is not great because of

the concentration in two-adult households. About 60 per cent of the households have two adults, about 10 per cent have more than three adults and about 1 per cent have more than five. Another result of this high concentration is that, unless the variable has a high correlation with the number of adults in the households, the difference between the weighted estimate and that in which each respondent has equal weight will be small; in case of small samples this difference may be negligible compared to the sampling error. For all attitudes thus tested in these studies the difference was inconsiderable compared with sampling error. However, in the comparison given in Table I there appear two differences of about two percentage points.

In order that the procedure may be applied and checked without great difficulty it is desirable to have a variable for ordering the members of the household; a variable than can be obtained by the interviewer objectively and easily. The age and sex of the members of the household were used for this purpose.

The Procedure

A "face sheet" is assigned by the sampling section to each sample dwelling unit; on each of these there are, in addition to the address of the dwelling, a form for listing the adult occupants, and a table of selection. At the time of the first contact with the household, the interviewer lists each adult separately on one of the six lines of the form; each is identified by entering in the first column his relationship to the head of the household (wife, son, brother, roomer, etc.). In the next two columns the interviewer records the sex and (if needed) the age of each adult. Following this the interviewer assigns a serial number to each adult: first the males are numbered in order of decreasing age, followed by the females in the same order. To assign these serial numbers it is necessary to obtain the ages of all adults only in that small portion of households in which there are two adults of the same sex and not connected by parent-child relationship.

Then the interviewer consults the table of selection; this table tells him the number of the adult to be interviewed. One of the six tables (A to F) is printed on each face sheet; each of the six tables is assigned to one-sixth of the sample addresses in a systematic manner.

Tables to select the person to interview:

Each of these tables was assigned to one-sixth of the sample addresses in the four surveys.

A	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	1	3	2	5	1	

B	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	2	1	3	4	2	
C	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	1	2	4	1	3	
D	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	2	3	1	2	4	
E	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	1	2	1	3	5	
F	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	2	1	4	3	6	

The slightly changed order of selection in the tables given below has certain advantages over that of the above tables:

The low numbers of selection are concentrated in tables A, B and C therefore the procedure will yield a male respondent in a great majority of the addresses to which these tables have been assigned. Evening calls are necessary to find at home most of the male respondents; the interviewer may concentrate his evening calls at these addresses. Conversely the interviewer may use his time during the day by calling at addresses to which tables D, E and F are assigned, with an increased chance of success.

The proper fractional representation of each adult is approximated more closely without the necessity of printing many more forms; the chances of selection are exact for all adults in households with 1, 2, 3, 4, and 6 adults. Because numbers above six are disallowed, there are one or two adults in a thousand (generally young females) who are not represented; there is a "compensation" for them in the over-representation of number five in the households with five adults.

Relative frequency of Use	Table Number	If no. of adults in household is					
		1	2	3	4	5	6 or more
		Select adult numbered					
1/6	A	1	1	1	1	1	1
1/12	B1	1	1	1	1	2	2
1/12	B2	1	1	1	2	2	2
1/6	C	1	1	2	2	3	3
1/6	D	1	2	2	3	4	4
1/12	E1	1	2	3	3	3	5
1/12	E2	1	2	3	4	5	5
1/6	F	1	2	3	4	5	6

It may be noted that the procedure can be modified easily to select a constant proportion, say half or one-third, of the adults. In that case, of course, each adult would have the same chance of selection, regardless of the size of the household. In some of the households more than one interview would be taken, in some no interviews at all. The above tables may be modified readily to apply the changed procedure.

Results: Checks against outside sources

The distributions of the respondents were checked with outside sources whenever we could obtain or adapt reliable data for valid comparison. They are presented in Tables I and II.

TABLE I

COMPARISON OF RESPONDENTS IN SAMPLE WITH CHECK DATA

	Check Data	Total for 4 Surveys N=2372		Data from the Surveys of:			
		Weighted	Unweighted	June 1946 N=585	Aug. 1946 N=592	Dec. 1946 N=570	April 1947 N=625
White	90.6	89.1	89.4	89	90	88	89
Native born	90.0	90.6	90.8	90	91	91	90
Age in years							
21–29	22.8	22.7	20.6	22	24	22	22
30–44	33.9	33.0	34.1	34	33	36	30
45–59	25.6	27.3	26.7	26	27	25	32
60 and over	17.7	16.1	17.8	17	15	17	15
NA		0.9	0.8	1	1	0	1
Education							
Not finished grammar	46.1	25.7	26.2	24	25	24	30
Finished grammar		19.1	19.3	17	17	22	20
Some high	17.4	17.1	17.4	18	19	18	14
Finished high	22.9	20.4	20.0	20	21	18	22
Some college	7.1	9.9	9.4	11	11	10	7
Finished college	5.0	7.0	6.9	9	6	7	6
NA	1.5	0.8	0.8	1	1	1	1

The check data are those deemed most nearly valid for comparison: the per cent white and the educational attainment are for April 1947 from Census Series P-20 No. 15, age is for July 1946 from Series P-S No. 19, and the per cent native born is the 1940 census figures for ages 16–64. In the data obtained from the survey results each respondent is weighted by the number of adults (from one to six) in his household. However, the column "unweighted", in which each respondent has unit weight, is included for purposes of comparison.

The appropriate sampling error (i.e. two standard errors) of the

estimates in Table I range from three to six percentage points for various items on each of the surveys, and from two to four percentage points for the four surveys combined. The estimates of color, nativity and age groups are in general agreement, with only one of the 30 estimates lying slightly beyond the given ranges.[7]

The data of the first three surveys showed an apparent upward bias in reports of college education. This was suspected to be a response bias to the single question on this topic. For the fourth survey three more questions were inserted to get more data on persons claiming 12 years of school or more; this involved additional work on only a third of the respondents. The fourth survey shows close agreement with the check data on college education, pointing to elimination of a response bias.[7]

Check of the sex ratio

TABLE II

PERCENTAGES OF MALES

Check Data: Census Estimate July, 1946	All Adults in Sample Households	Respondents in Sample Weighted by Number in Household				
		4 Surveys Combined	June 1946	August 1946	December 1946	April 1947
48.2	47.9	46.8	45	46	45	52

The Bureau of Census estimate is for the civilian non-institutional adult population for July, 1946 (series P-S No. 19). The second figure is an internal check, obtained by tabulating all adults listed as living in sample households.

Males appeared to be underrepresented among the respondents of the first three surveys. Although the difference was small, its presence in three surveys pointed to possible occasional deviation from rigorous procedure in the field. It may be noted that a 3 per cent bias in the sex ratio would require an unlikely 33 per cent sex difference in an attitude studied in order to produce a 1 per cent difference in survey results; a difference within the limits of accuracy of these surveys.

Tabulations of all adults in the sample households, as listed on the face sheet, gives close agreement with the check data. Continued investigation of the problem points to two sources of this small bias, both due to the fact that males are more difficult to find at home even with repeated call-backs: overrepresentation of males among the non-

[7] These statements are also borne out by results on three more national surveys in which the same procedure was used.

responses, and an occasional substitution on part of a few interviewers.

Two other small sources of error present in the first three surveys were eliminated by minor changes in the procedure.

Conclusion

The described procedure of selection within the household gave results that were satisfactory within the demands of the survey objectives. While there were occasional departures from correct procedure in the field, the procedure is such that extensive control over its field application, hence improvement, is feasible. It must be emphasized, however, that a practical sampling procedure is not an automatic device. For success it depends on a field force having both the training and the morale necessary for correct application.

BENEFICIARY STATISTICS UNDER THE OLD-AGE AND SURVIVORS INSURANCE PROGRAM AND SOME POSSIBLE DEMOGRAPHIC STUDIES BASED ON THESE DATA*

ROBERT J. MYERS
Chief Actuary, Social Security Administration, Washington, D. C.

The old-age and survivors insurance program, under the Social Security Act, is a nationwide social insurance system covering most employment in industry and commerce. It is possible that in the near future this coverage will be extended to virtually all employment in the country since almost all interested organizations are in favor of this. From the extensive scope of the present limited coverage and also of the anticipated, practically universal future one, there naturally will emerge over the years a very large number of beneficiaries, both retirement and survivor cases. This paper will first present a limited amount of background regarding the present program. There will then be set forth the sources and types of data now being presented or available; finally there will be described some demographic studies that have been made, as well as others which are possible from these data.

Coverage of the Program

THE EMPLOYMENT coverage of the present system includes substantially all wage and salary workers in industry and commerce except those engaged in railroad employment for whom a separate system has been set up. The major employment categories excluded from coverage are agricultural, domestic, governmental, non-profit, and railroad, as well as self-employment. Table 1 presents certain relevant data as to the extent of coverage under the system.

Benefit Provisions and Definitions

Monthly retirement benefits are available under the OASI program for "fully insured"[1] workers age 65 and over, and for their wives 65 and over and their dependent children under 18. These payments are not

---

* Presented at the 108th Annual Meeting of the American Statistical Association in Cleveland, Ohio, December 27–29, 1948.

[1] Roughly, defined as having had covered employment of a significant amount ($50 or more) in at least half of the calendar quarters since 1936, or attainment of age 21 if later, and before attainment of age 65 or death. Employment before age 21 and after 65 is credited toward meeting the requirement even though it does not serve to increase the requirement. In no case is there required more than 40 such quarters.

TABLE 1

COVERAGE OF THE OLD-AGE AND SURVIVORS INSURANCE PROGRAM

Description of Coverage	Number involved
1. Persons in covered employment during average week in 1947	34 million
2. Persons in covered employment some time or other during 1947	49 million
3. Persons, at end of 1947, who were at some time in covered employment during 11 years of the program	77 million
a. Number, in this category, who had earned sufficient wage credits to be insured	43 million
Average cumulative wage credits for this group	$11,000
b. Number with insufficient wage credits to be insured	34 million
Average cumulative wage credits for this group	$1,300
4. Equivalent amount of insurance in force under survivor provisions of the OASI system at end of 1947	$75 billion

made, however, if the individual earns $15 or more per month in covered employment. As a result, many individuals do not file claim upon attainment of age 65, since they continue at work. On the other hand, some persons file a retirement claim but return to work, either permanently or sporadically. Beneficiaries who actually receive monthly payments are said to be in "current payment status." When there are added those who have filed but are not currently receiving payment because of covered employment, the total is termed "benefits in force." At the end of 1948, there were 1.97 million persons age 65 and over who were fully insured, but only 53 per cent were in current payment status. Those who had filed a claim but were not in current payment status represented 9 per cent, while the remaining 38 per cent never had filed; thus 62 per cent of the benefits were in force.

Survivor benefits are available[2] to widows when they are age 65 or

[2] All survivor benefits are payable if the deceased worker was "fully insured"; in addition, even though "fully" insured status was not present, monthly benefits are payable to orphans and widowed mothers where "currently insured" status (roughly, covered employment of $50 or more in half of the quarters of the last 3 years) existed.

over, or when they have 1 or more dependent children under 18 in their care. Monthly benefits are available for children under 18 who are orphaned. In certain cases, there are monthly benefits for aged parents of deceased workers, if there was no surviving widow or child under 18. These survivor benefits are subject to the same concepts of current payment status and benefits in force.

Table 2 gives certain summary data on beneficiaries. It may be

TABLE 2

BENEFICIARIES OF THE OLD-AGE AND SURVIVORS INSURANCE PROGRAM IN CURRENT PAYMENT STATUS, DECEMBER 1948

Category	Number (in thousands)	Average Monthly Benefit*
A. Retirement		
Primary	1048	$25
Wife (supplementary)	321	13
Child (supplementary)	25	12
B. Survivor		
Widow, old-age	210	21
Widow, with children	142	21
Child	556	13
Parent	12	14
Grand Total	2314	†

* Rounded to the nearest dollar.
† Not computed because not significant.

noted that the 2.3 million beneficiaries are receiving payments at an annual rate of about $550 million. It has been recommended by the Social Security Administration and the Advisory Council of the Senate Finance Committee (Senate Document 208, 80th Congress, 2nd Session) that the benefit level be raised and additional beneficiary categories be added.

Another concept involved is the distinction between "initial" and "subsequent" entitlements as represented in the new claim awards of a particular year. An "initial" entitlement is the first claim on any particular case, such as a worker retiring, a wife's benefit claim filed when the husband retires, or a widow's benefit claim when she is over 65 or has dependent children when her husband dies. A "subsequent" entitlement is a claim which would not have resulted in monthly payments

at the time of the insured's death, as for instance a wife's benefit claim when she attains 65 after her husband retired or a widow's benefit claim when she attains 65 after her husband's death.

In contrast with temporary suspension of benefits, as for instance because of covered employment, there are a number of reasons for permanent termination of claims, such as death, remarriage of widows, and marriage or attainment of age 18 of child beneficiaries.

Sources of Data

Beneficiary data are published in a number of official publications of the Social Security Administration. The *Annual Report* carries only general summary data, such as number and amount of benefits in current payment status and total payments certified, allocated by state and also given for the entire country by category of beneficiary.

The *Social Security Yearbook* presents a large volume of statistics on a calendar-year basis. The most basic beneficiary data are the numbers and monthly amounts of benefits awarded, in current payment status, and in force by beneficiary type for various periods (also carried forward monthly in the *Social Security Bulletin*). There are given data on benefits suspended and on benefits terminated, classified according to reason. Data for the current year are shown for new awards of the year, benefits in force, and benefits in current payment status by age, sex, and race, as to both number and amount. Although the data are summarized into fairly broad age groups (generally quinquennial) because of desired economy in publication, the figures are actually available by single years of age. This, in many instances, is necessary for proper analysis, especially for retired persons where there may be a great difference between those who retire promptly at age 65 and those who defer retirement. Beneficiary data are also available to indicate in summary form the family composition of the beneficiary groups, both for new awards in the year and for benefits in current payment status at the end of the year.

The aggregate data on benefits are an actual count, whereas the other data giving details by age and family classification are based on a 20 per cent sample.

In addition to published periodical data, there is prepared each year a volume of substantive claims statistics which shows in considerable detail the claims awards (on a 20 per cent sample basis) by age of retired or deceased wage earner and type of benefits available.

Once each year there is obtained a count of the number of beneficiaries actually receiving payments in each county in the United

States. These data are subdivided only by type of benefit and not by age, sex, or other family characteristic and are used not only for publicity purposes, but also for contrast with the recipient-load under the various public assistance programs.

At irregular intervals, sample surveys on an interview basis are made to study various phases of beneficiary living conditions and resources. These surveys, although being based on the interview approach and subject to some limitations, are stratified samples of beneficiaries from the claims files. The results of these studies are published in the *Social Security Bulletin*, or in special releases. Also from time to time, there are issued various *Analytical Notes* which sometimes relate to special studies and tabulations dealing with different characteristics of beneficiaries. In addition, some of the *Actuarial Studies*, published intermittently, analyze beneficiary data and present future estimates thereof.

Adaptability of OASI Data for Demographic Studies

There are many demographic studies which may be made from the basic data collected under the OASI program. Some of these raw data now exist in tabulations and in analyses which are under way. In other cases, different types of coding and tabulating are necessary. For instance, in the past, deaths or other events such as remarriage have not been allocated by the date of occurrence, but rather classification was by date of administrative action, either original filing or final award. Complications, too, arise because of the several concepts described previously, namely, in force status versus current payment status and initial entitlement versus subsequent entitlement. These various complications can probably be resolved or taken into account by appropriate approximation or adjustment.

The various basic data which are tabulated and analyzed are necessary in the actual operation of the program. This is particularly the case in regard to actuarial cost estimates and other analyses for policy and planning, and also from the administrative standpoint in estimating work loads for the near future.

Mortality Studies

Certain crude mortality studies have been made. For primary beneficiaries (retired workers), it has been found that as contrasted with general population experience, the mortality has been somewhat higher for men (by roughly 5 to 15 per cent), but somewhat lower for women (by about 15 to 20 per cent).

It appears likely that mortality for male primary beneficiaries is

relatively unfavorable because with present high employment conditions, those who have actually retired, in most instances have done so only because of disability. This argument is further strengthened by the fact that those who retired promptly at age 65 and those who retired in 1940, the first year in which benefits were available, tended to have higher mortality than other retirants.

The mortality for female primary beneficiaries, despite the factors previously mentioned as relating to male mortality, may have been lower than normal because women who had been working near age 65 long enough to be insured might, on the whole, be superior lives. For both sexes, the death rates appear to have been independent of the size of benefits, which in turn reflects economic status.

Considering other types of beneficiaries, studies indicate that both wife and widow beneficiaries, all of whom are age 65 or over, had relatively low mortality as contrasted with general female mortality, namely, by an amount of 10 to 15 per cent relatively. The mortality among child beneficiaries, the vast majority of whom are paternal orphans, is about 10 per cent higher than would be expected from general population experience. At the same time, the mortality of the widowed mother beneficiaries is quite close to the general population experience.

A recent study made primarily to study remarriage experience has indicated that for white widowed mothers, the mortality was perhaps 5 per cent below that "expected," but for nonwhite widowed mothers there was lower mortality by about 15 to 20 per cent. The relatively favorable nonwhite mortality may occur because those nonwhite workers who are regularly in covered employment sufficiently to be insured probably represent those of above average economic status, and therefore the health conditions of their family would tend to be above average.

Unlike studies based on census data, the particular individual beneficiaries under OASI will be under observation over many years so that there is considerably more accuracy as to age reporting, both as to the living and as to the deaths. Eventually, there will be a large body of data which will yield information on that great "unknown" of mortality investigations, namely, the true experience at the older ages (say, 85 and over).

Also, from the OASI data it will be possible to examine mortality by marital status. Certain Census studies in the past have shown that married persons have somewhat lower than average mortality and that widowed persons have relatively high mortality. Preliminary studies

made from OASI data indicate that although married beneficiaries, both the retired workers and their spouses, have somewhat favorable mortality, there is no indication of very unfavorable mortality among widows, whether beyond age 65 or not.

By having the same group of lives continuously under observation, it will be possible to construct "generation mortality tables," which will give valuable indications as to the mortality of a specific group of individuals traced through from age 65, or slightly older, to eventual death. This contrasts with Census methods where it is only possible to study mortality by age for different groups of persons in a given year or short period of years; for instance, the experience at ages 80–84 will relate to different individuals than that at ages 65–69. The objection may here be raised as to amalgamating data which are not entirely homogeneous since they do not relate to a closed group of persons.

Remarriage Studies

The only available analyzed information on remarriage experience in the United States has been the American Remarriage Table which relates to a relatively small set of data arising from workmen's compensation claims under private insurance carriers.[3] Under the OASI system, there is a great body of experience since the number of new widows created each year is close to 100,000. For a study of this type it is necessary to consider both age and duration of widowhood.

One crude study that has been made in regard to the remarriage experience under the OASI system indicates that the actual remarriage rates are considerably higher than those of the American Remarriage Table, apparently by from 50 to 75 per cent. Several explanations seem likely. Since 1940 marriage rates have been quite high so that perhaps remarriage rates too have been high. Also the level of benefits under the OASI system is relatively low as compared to that under workmen's compensation claims so that there is less financial deterrent to remarriage.

At the lower end of the age scale, it is possible to obtain information as to mortality and marriage of children under 18 since these are reasons for termination of child's benefits.

Other Studies

If permanent and total disability benefits are added to the OASI program, as has been proposed, there will eventually be much data on

[3] Roeber, W. F. and Marshall, R. M., "An American Remarriage Table." *Proceedings of the Casualty Actuarial Society*, Vol. XIX.

disability incidence and termination rates. This has been a subject on which there is very little material in the United States other than the experience under life insurance policies and certain pension plans for governmental employees. Some students feel that neither of these sources of data nor the experience of foreign social insurance systems is indicative of the possible benefit experience which might arise for such a program in this country.

The beneficiary data under the OASI system can also yield information on family composition. Among the various data which could be obtained are the following: relative ages of husbands and wives, marital proportions, proportion of persons with children and numbers of such children, etc. In all instances, the analysis could be carried forward by age and race.

In the preceding discussion, the various demographic studies have been discussed on a nation-wide basis since this is the nature of the OASI program. However, the analyses could be subdivided by regions or even by states, but then there would be considerable administrative difficulty because the records are not set up to show existing residence combined with other demographic statistics. Likewise, any studies of migration of beneficiaries, although theoretically possible, would not be administratively feasible.

In addition, the analyses could in all instances be made by economic level since benefits payable are related to past covered wages; this would be especially valuable if coverage were extended to all employment, but under present limited coverage the results would not be too significant.

Although it would be highly desirable to make mortality and other investigations according to occupation, a subject on which there is available all too little information, it would appear that this would not be too feasible from the OASI data. For one thing, the records which are obtained through the working lifetime do not indicate occupation, although they do show the industries in which the worker was employed. However, because of the dynamic character of the labor market, many individuals shift frequently from one type of industry to another so that even if the lifetime data were available for each worker at time of retirement, there would be considerable difficulty not only theoretically, but even more important administratively, in classifying the "normal" industry.

As is apparent from the wide scope of the program, there is a vast amount of statistical data in the files which has not been tabulated or collected. The manpower limitations during the war years, of course,

had an appreciable effect on studies that might desirably have been made. Since the war, there have been both manpower and budgetary limitations.

As contrasted with the various mortality and other types of investigation that insurance companies do, the Social Security Administration has found it possible to do only a limited amount of actuarial and statistical investigations. In part, this might be explained by the fact that insurance companies must keep a close surveillance of their experience in order to be able to charge proper premium rates and to allocate dividends equitably.

On the other hand, under a governmental system, there are not such concepts of individual equity. Rather the benefit schedules are set by law and are payable accordingly, with only a general examination being required as to the immediate and long-range over-all financial balance. However, it is hoped that in the near future more extensive tabulation and collection of data will be possible so that some of the demographic analyses indicated previously may be made.

BY-PRODUCT DATA AND FORECASTING IN UNEMPLOYMENT INSURANCE*

NATHAN MORRISON
Division of Placement and Unemployment Insurance
State of New York

The unemployment insurance program has been operating in this country since the beginning of 1938. There have been many growing pains during these first eleven years. The operating problems involved in setting up this vast program have left little time thus far for careful consideration of the statistical data that are automatically produced in the course of its operations.

These data, now and in the future, can be a fruitful source of information on changes in economic activities. The comprehensive nature of these data, covering every city and county, giving information by detailed sub-industry, and yielding statistics for each week and month, makes it possible to test various theories of the business cycle, and may, as a result of careful analysis, reveal interrelationships among economic variables that are now unsuspected.

The data described in this paper are collected by all the state unemployment insurance agencies as part of their operations. However, some states publish only the statistics which are needed for administrative and operating purposes. Undoubtedly, more would be published if the existence of these valuable materials became generally known.

BY-PRODUCT DATA

WHAT DOES the unemployment insurance system offer in the way of by-product data? Essentially, it provides detailed geographical and industrial data on employment and payrolls in firms covered by the state unemployment insurance laws, and on unemployed persons claiming benefits.

The coverage of the unemployment insurance program is not yet complete. About 2 out of every 3 workers in the country are now included. In manufacturing, however, about 98 per cent are covered. The chief groups excluded are government employees, agricultural workers, self-employed persons, workers in some small firms (the state laws vary in this respect), employees of non-profit organizations, and domestic

* Presented at the 108th Annual Meeting of the American Statistical Association in Cleveland, Ohio, December 27, 1948.

workers. (Railroad workers have a separate unemployment insurance system.) Federal and state legislation to include practically all of these groups may be confidently expected within the next few years.

For many industries and geographic areas, the present laws represent substantially complete coverage. In terms of long-range studies, work can be planned on the safe assumption that full coverage will be reached in the near future.

A concrete example is perhaps the best way to indicate the data available and some of the ways these data can be used.

For every city and county in New York State, tabulations are available showing employment and payrolls by detailed industry group. The industry classification uses a 4-digit code that gives about 600 different sub-industry groups. The employment data show the number of workers during the mid-week of each month. The payrolls are on a quarterly basis. Since 1937, all employers with 4 or more employees (excluding the groups mentioned earlier) have been submitting such employment and payroll reports together with their tax contributions on a regular routine basis, as a normal activity of the unemployment insurance system. Actually, these reports contain more than the summary data on employment and payrolls. The name, social security number, and earnings of each individual worker are listed and submitted also each quarter. The reports are needed in order to collect the tax and to provide the individual earnings data used to determine benefit rights when an unemployed person files a claim for unemployment insurance.

These reports are mandatory under the law with severe penalties for failure to comply. The statistical data on employment and payrolls are collected as a by-product of necessary operations and are thus obtained at a minimum cost.

Another operating feature is the maintenance of lists, card files, and addressograph plate files, of all covered firms. Thus if a sample of firms is needed, the universe from which it is to be selected is readily available.

The usefulness of these data on establishments, employment and payrolls, by industry and geographic area, for market analysis, industry studies, labor market studies, and other related purposes is obvious. For study of changes in economic activity, Arthur F. Burns[1] has suggested the preparation of tables showing not only the total change in employment each month by industry, but also the number of firms increasing their work force and the number of firms dropping workers.

[1] Arthur F. Burns—"Twenty-sixth Annual Report of the National Bureau of Economic Research," June 1946, p. 22.

Another type of analysis that may throw some light on economic changes is the study of the employment changes in each industry by size of firm. Some studies in certain New York industries have indicated that the largest and smallest firms have more stable employment than the medium-sized firms. New York State has also analyzed employment patterns in the millinery and dress industries in separate groups according to the price of the product, and in the canning industry by the crop being canned.

The collection of these employment and payroll data has involved problems of industrial classification and other troublesome details that are to be encountered in any large-scale undertaking. Such problems will always be with us. They need not concern the analysts who will use these data. One particular problem may be mentioned. At present, all employment and payroll reports submitted to the state agencies, and the monthly sample of establishment reports which are handled in cooperation by the Bureau of Labor Statistics and the state agencies, both cover the mid-week of the month (i.e., the week ending nearest the 15th of each month). The achievement of this uniformity is an obvious step, yet it required considerable effort and discussion.

The unemployment insurance agencies also provide data on unemployment as a by-product of their operations. The 48 states and Alaska, Hawaii, and the District of Columbia all have individual laws with varying eligibility requirements for benefits. These minor differences have been given exaggerated importance by some persons who are not familiar with the practical situation. Essentially, the unemployment insurance laws define an unemployed person as anyone who is not working, is seeking work, and is willing and able to take a job if offered one. This definition is simple and unambiguous, and is in agreement with the average person's idea of unemployment. An unemployed worker, in order to obtain insurance benefits, must report in person at the local unemployment insurance office nearest his residence, and must fill out forms and answer verbal questions which are designed to obtain information concerning his occupation, last employer, reason for loss of employment and his willingness and ability to take a job. This direct, personal contact with the unemployed worker is maintained, for he must report to the office once each week as long as he continues to be unemployed and seeking benefits.

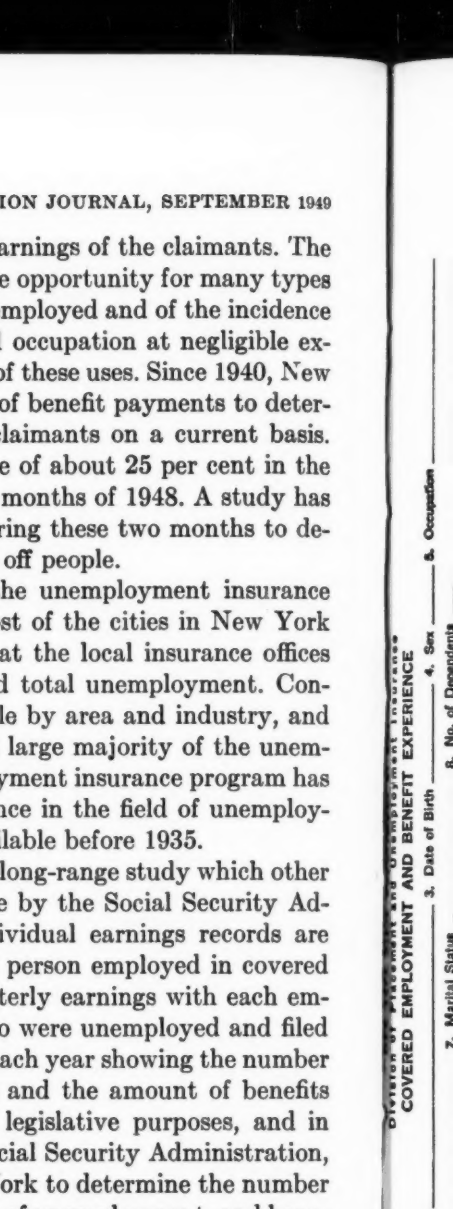
Operating reports prepared weekly by each local office give the number of persons reporting each week by type of claim, i.e., first claim and continued claim. Central offices, which usually prepare and mail out the benefit checks, provide as part of their operations, additional data on

the previous industrial attachment and earnings of the claimants. The existence of these operating data offers the opportunity for many types of studies of the characteristics of the unemployed and of the incidence of unemployment by industry, area, and occupation at negligible expense. Two examples will illustrate some of these uses. Since 1940, New York has made a monthly sample study of benefit payments to determine the industrial distribution of the claimants on a current basis. More recently, there has been an increase of about 25 per cent in the number of claimants during the last two months of 1948. A study has been made of all first claims received during these two months to determine the industries which were laying off people.

The present incomplete coverage of the unemployment insurance laws is only of minor significance. In most of the cities in New York State, the persons now seeking benefits at the local insurance offices represent about 80 per cent of estimated total unemployment. Considering the detailed information available by area and industry, and the personal, repeated contacts with this large majority of the unemployed group, it is clear that the unemployment insurance program has provided as a by-product, a major advance in the field of unemployment statistics, far beyond anything available before 1935.

New York State has been working on a long-range study which other states have also been urged to undertake by the Social Security Administration.[2] Beginning with 1937, individual earnings records are available for each year to date for every person employed in covered establishments. These records show quarterly earnings with each employer. In addition, for those persons who were unemployed and filed claims for benefits, data are available for each year showing the number of weeks of compensable unemployment and the amount of benefits received. For many administrative and legislative purposes, and in order to supply data requested by the Social Security Administration, an annual study has been made in New York to determine the number of benefit claimants by industry, duration of unemployment, and benefit rate. By taking as a sample the approximately 10 per cent group of persons whose Social Security account numbers end in the block 2,000 to 2,999 (referred to as the "2,000 block"), and by entering the information on individual record cards which have space for 14 years of employment and unemployment data, it has been possible to obtain, as a

[2] A comparable long-range study of employment data available under the Old Age Insurance Program is described in "*The Continuous Work History Sample under Old Age and Survivors Insurance*" by Jacob Perlman and Benjamin Mandel, Social Security Bulletin, February 1944. New York's study has been developed in collaboration with the staff of the Social Security Administration.



Division of Placement and Unemployment Insurance

COVERED EMPLOYMENT AND BENEFIT EXPERIENCE

1. S. S. No. _____ 2. Name _____ 3. Date of Birth _____ 4. Sex _____ 5. Occupation _____

6. Home Address _____ 7. Marital Status _____ 8. No. of Dependents _____

1937			1938			1939			1940			1941			1942			1943		
Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.
Tot.																				
H.Q.																				

	1938-39 BEN. YR.	1939-40 BEN. YR.	1940-41 BEN. YR.	1941-42 BEN. YR.	1942-43 BEN. YR.	1943-44 BEN. YR.	1944-45 BEN. YR.
Did not apply							
Eligible claim							
Ineligible claim							
Local office							
Date of orig. claim							
Date of last check							
Time span of claim							
Spells of unempl.							
Benefit rate							
No. of checks							
Total benefits							

RS 647 (4-46) 4-21-48-50M (8-OMO-48)

FIGURE I. FORM CARD FOR EMPLOYMENT AND BENEFIT EXPERIENCE.

by-product of these required annual studies, a continuous record of the employment pattern of a large sample of New York workers over a period of many years. The detailed analysis of these materials has thus far been limited by recurring budget difficulties resulting from inadequate financing of unemployment insurance administration by Congress. However, the data are produced each year as by-products of the operations of the unemployment insurance system, and are being preserved for analysis whenever adequate staff becomes available. This long-range study of employment and unemployment by industry and area since 1937, which can be made in New York and the other states, should provide valuable data for both theoretical and practical work of students of the business cycle, business and market analysis, and economists and statisticians in general. (See figure 1.)

These materials have also been used to a considerable extent in forecasting work. This aspect can be touched on only briefly in this paper. The forecasting is done for two major purposes. Unemployment insurance benefits are paid out of a trust fund which in New York has totalled about $1 billion since 1945. The solvency of this trust fund is of course a major concern. Each year, all legislative proposals for changes in the benefit formula or in the taxing formula are analyzed and estimates of the effect on the solvency of the fund are prepared under various assumptions as to the level and movement of employment and payrolls during a 5 to 7 year period in the future. The effect of the end of the war was the subject of a comprehensive report which used all available materials on employment and unemployment by industry and area as the background. The second major purpose is to estimate claim loads for budgetary and planning use. Administrative funds are provided by grants from Congress. A year in advance of the beginning of each fiscal year, estimates of unemployment insurance claims loads expected during the fiscal year must be submitted by each state. In addition, forecasts are prepared for each semi-annual period about four months before the beginning of each period; and finally monthly forecasts are prepared for each local office about two weeks in advance in order that local office staffs may be shifted to meet the load most efficiently. New York's experience in eleven years of operation may be of some interest and perhaps can be presented in detail in another paper.

CONCEPTS OF EMPLOYMENT AND UNEMPLOYMENT

Early in 1948, the International Labour Office issued a report on "Employment, Unemployment, and Labour Force Statistics" which is

subtitled "A Study of Methods." This report was prepared for the Sixth International Conference of Labour Statisticians held in Montreal in August 1947.

This I.L.O. report comments in some detail on the methods being used in this country and in other nations to collect data on employment and unemployment and suggests a broad approach which is essential if the available statistical resources are to be used efficiently.

The collection of data on employment and unemployment in this country is being done for many purposes, and by many different agencies. However, only the unemployment insurance system provides detailed geographical and industrial information on both employment and unemployment entirely as a by-product of its routine operations. It is for this practical reason that it is likely that with universal coverage, the unemployment insurance program will be generally accepted as the primary source of this type of statistics. It should be noted here that as far back as 1937 the Committee on Government Statistics and Information Services which was jointly sponsored by the American Statistical Association and the Social Science Research Council, stated in its report on "Government Statistics,"[3] "the present voluntary system of payroll and employment reporting may be rendered obsolete with the advent of unemployment compensation laws, which promise to provide comprehensive returns as a by-product of administration."

There will of course continue to be need for some of the other methods now being used. Again the reason is a practical one. The problems that statisticians are concerned with in economic statistics in general and in studying changes in the level of employment in particular are complex. The problems will be solved, but as Professor Neyman remarked in 1937 in discussing the analysis of economic time series, "it will not be done today or tomorrow." It is important to recognize that all available tools and resources are needed if significant progress is to be made in this field of research. In a paper published 21 years ago, Professor Harold Hotelling[4] pointed out that the Law of Gravitation could not have been established as it was by Newton if the sun were smaller and closer in size to some of the planets. If the sun were smaller, the family of planets would exhibit some of the complex movements of democratic societies and statistical methods not known in the 17th century would have been necessary to determine the Law of Gravitation.

The problems of employment and unemployment are important to

---

[3] "Government Statistics"—Bulletin 26, Social Science Research Council, April 1937, p. 92.

[4] Harold Hotelling—"Differential Equations Subject to Error and Population Estimates," *Journal of the American Statistical Association*. Sept. 1927, p. 287.

many groups and agencies in this country. The Joint Congressional Committee on the Economic Report in its recent report on "Current Gaps in Our Statistical Knowledge" refers to the need for extension of employment and unemployment statistics. The Committee states:

> "Such statistics are among the most important indicators of the economic situation. In this area, the Census Bureau's Current Population Survey is particularly valuable. The chief weakness of this monthly enumerative survey is that it fails to show employment and unemployment on a geographical basis, and that the sample is too small to provide reliable data on occupational and other characteristics of the unemployed. . . . The need for geographic detail can be met in part by annual surveys of individual metropolitan areas . . . Annual surveys of employment and unemployment for the major urban areas do not provide sufficiently current data in a period of rapid change. For this purpose primary emphasis must be placed on employment data reported by employers. State data of this sort are compiled by state employment security agencies and state labor departments in cooperation with the Bureau of Labor Statistics which also compiles the corresponding national series, but area series are maintained by only a few state agencies."

There is no reference at all to the unemployment insurance data available on claims filed by unemployed persons which every state agency collects as a by-product of its operations and which are actually reported to the Social Security Administration in Washington each month for each local office in every state.

The reference to the employment data reported for each month by all covered employers fails to emphasize the important fact that the data are available as a by-product of operations, and that the additional money needed for analysis and publication is but a tiny fraction of the cost of obtaining such detailed data on a current basis by any agency not conducting a vast program like the unemployment insurance system.

It is not possible in this paper to discuss the Census Bureau Household Sample in any detail. Obtaining information by visiting a sample of households has some advantages and some important defects. This tool of research can be a useful supplement, at present, to the unemployment insurance by-product statistics. When universal coverage is attained, there will be little need for a household sample, except perhaps for some special purposes.

There is an important point concerning the definitions of employment and unemployment. The Census Bureau uses a definition of unemployment which is not in accord with the concept used in the state laws described above. The definition suggested by the International Labour Office report is similar to the one used by the unemployment

insurance laws, namely that "the unemployed should include all persons seeking work on a given day who are not employed but are able to take a job if offered one."

The I.L.O. report also comments on the Census Bureau insistence on including as unemployed only those who have been without work for at least the Monday through Saturday period in the enumeration week. The I.L.O. report estimates that unemployment was thus understated about 7 per cent during the period June 1941–May 1942.

It would be a major step forward if the Census Bureau would bring its concepts and definitions more in line with the concept of unemployment which underlies the laws of the 51 unemployment insurance agencies. Whatever technical difficulties stand in the way, they must be overcome if the Census Bureau work in this field is to play its proper role as a supplement to the by-product data.

The I.L.O. report has some pertinent comments on the use of unemployment insurance statistics as a measure of unemployment. The limitations it mentions are of some significance at present. However, with universal coverage and a probable extension of duration of benefits to 26 weeks in every state, these limitations will no longer be significant.

One quotation from the I.L.O. report will serve to summarize some of the views expressed in this paper:

> "Social insurance statistics of unemployment are in a very real sense cost free, being by-products of the operation of a system installed for other than statistical purposes. It therefore becomes possible to expand statistical coverage and derive additional estimates (as well as to secure the basic estimates) at a remarkably low cost assignable to statistical purposes. Furthermore, the continuing contact of a social insurance system with individual workers makes it possible to conduct a variety of special studies on unemployment problems at low cost and with little inconvenience to the employee. The possibilities of using social insurance data for the study of unemployment problems have only begun to be explored."

The rich mine of statistical data available as a by-product of our unemployment insurance system if carefully analyzed may very well lead to significant progress in economic theory and to a better understanding of the movements of the business cycle. Analysts and research workers in this field should examine the data already available, and should encourage the analysis and publication of data which are being collected in all the states, but which, in some states, may merely be stored away or filed for lack of funds and outside interest.

STATISTICAL REQUIREMENTS FOR ECONOMIC MOBILIZATION

RALPH J. WATKINS
Office of Plans and Programs
National Security Resources Board

IT IS IMPORTANT at the outset to make clear the distinction between mobilization and mobilization planning. In dictionary terms, mobilization is defined as the "act of assembling, equipping, and preparing military and naval forces for active hostilities." It is scarcely necessary to say that the American nation is not engaged in mobilization; nor are we preparing for mobilization. Mobilization planning is quite another matter, as I shall try to show presently.

The National Security Act of 1947 established the National Security Resources Board and assigned to that Board the statutory function of advising the president concerning the coordination of military, industrial, and civilian mobilization, including advice to the President on certain specific matters having to do with effective mobilization of resources in the event of war and with certain economic readiness measures against the contingency of war. Thus these functions make of the National Security Resources Board an economic mobilization planning agency set up to advise the President. It is important to note that the Board has no operating functions in the governmental sense of the term; its duty is to advise the President.

Economic mobilization planning may be defined as the process of estimating the requirements or needs of war; of appraising the resources or means which would be available for meeting those needs; of measuring deficiencies revealed by the comparison of needs with means; and of determining the steps necessary to balance needs with means—all to the end that there may be available well-articulated plans for mobilizing the resources of the nation in the event of war. It is important to stress the words "in the event of war." We are not engaged in planning "for war"; rather, we are planning against the contingency of war.

The philosophy of mobilization planning, both military and economic, rests on the premise that a state of preparedness is one of the means of lessening the likelihood of an aggressive attack against the nation and at the same time one of the means of increasing the likelihood of winning a war, if the nation is forced into war. In the un-

certain world in which we live, we can with prudence do no less than to take appropriate steps to improve our economic readiness position against the contingency of war and to lay plans for the rapid and effective mobilization of our resources in the event of war.

It will be recognized at once that the definition of economic mobilization planning which I have outlined is one which in many fields—not all fields by any means—lends itself to translation in statistical terms.

For example, one of the specific statutory functions assigned to the National Security Resources Board is to advise the President on "the relationship between potential supplies of, and potential requirements for, manpower, resources, and productive facilities in time of war." We interpret that provision to mean that this advice must be given in time of peace against the contingency of war, as well as in time of war. The basic and fundamental tool in a program of economic mobilization planning consists of detailed balance sheets of resources and requirements—above all for key raw materials but also for major end products, critical components, fuels, electric power, transportation, and manpower.

In time of peace, such balance sheets provide the factual basis for advice to the President from time to time as to the steps that should be taken to strengthen our economic readiness position or to lessen our economic vulnerability against the contingency of war. In time of war, this balance sheet process would provide the factual basis for operations by the war agencies.

What are we doing to meet this fundamental need for resources—requirements balance sheets? We have under way three basic programs:

1st. Quick and approximate estimates of requirements to test the economic feasibility of strategic plans.

2nd. Balance sheets for Fiscal 1950 of key resources and security program requirements.

3rd. Detailed estimates of resources and mobilization requirements on a systematic and continuing basis.

These three programs will be outlined briefly.

First—*Quick and approximate estimates of mobilization requirements to test the economic feasibility of strategic plans.* The Munitions Board in the National Military Establishment is now developing quick estimates—or flash estimates as they call them—of military requirements under the assumptions outlined in the strategic plans. These will embrace the needs of the Military Establishment for the key materials of steel, copper, aluminum, and petroleum; for construction; for many key

items of military equipment; and for manpower. When completed, these estimates will be submitted to the National Security Resources Board. Concurrently with this work, the National Security Resources Board is securing comparable quick estimates of mobilization requirements from the other security agencies—the Atomic Energy Commission and the United States Maritime Commission. In cooperation with the Departments of Commerce, Interior, State, and Labor, we are also developing quick statistical estimates of the mobilization requirements of the civilian economy. These last named civilian mobilization needs are being estimated initially on the basis of the civilian economy as it existed at the peak of our war effort in World War II, with, of course, appropriate adjustments for changes in population, changes in inventories, and changes in patterns of consumption. We will then add these estimates to those supplied by the Munitions Board to secure a measure of estimated *total* mobilization requirements.

We will next compare these mobilization needs with the estimates of available resources, which are being prepared also in cooperation with the Departments of Commerce, Interior, State, and Labor. On the basis of that balancing of mobilization needs against means, we will be able to say to the National Military Establishment whether the strategic plans can be encompassed within the limits of our resources, and if not, why, and in what particulars.

Second—*Balance sheets for Fiscal 1950 of key resources and security program requirements.* At the request of the President, the Board has undertaken, with the cooperation of 21 federal departments and agencies, to appraise the resources available to the nation in relation to the requirements in Fiscal 1950 of the several national security programs—current and anticipated. That survey will include a wide range of balance sheets covering the major strategic and critical materials, fuels, power, transportation, manpower, and the key manufacturing industries of the nation. These balance sheets will serve as the basis for advice to the President on questions of national security policy. For example, they will indicate: (1) to what extent our physical resources may set limits to the national security programs; (2) to what extent these programs can be accomplished without controls; (3) to what extent controls would be required to assure completion of these programs; or (4) to what extent controls would be required to protect the economy from price spirals in the raw material markets. Likewise, these balance sheets will afford a factual basis for economic readiness measures aimed at increasing our resources which the Board may see fit to recommend to the President.

Third—*Detailed estimates of resources and mobilization requirements*

on a systematic and continuing basis. Coming back to the assumption that the strategic plans meet the tests of economic feasibility which I have described: On the basis of the assumptions outlined in those plans, the National Security Resources Board must assemble from the several federal departments and agencies on a continuing and systematic basis estimates of key and major resources and key and major mobilization requirements, both military and civilian. The extent to which this statistical planning process can be projected currently remains to be determined. In consequence, I can merely outline in general terms my own views. The Munitions Board, with the cooperation of the Service, will estimate detailed military mobilization requirements for submission to the National Security Resources Board. The Board must look to the appropriate civilian deparments and agencies for the estimation of civilian mobilization requirements and likewise for estimates of available resources, as their continuing areas of responsibility in the task of economic mobilization planning.

The staff of the National Security Resources Board is now engaged in the preparation of requirements manuals. These manuals will outline the necessary assumptions, methods, procedures, and forms. Out of this process could be developed balance sheets covering major and critical needs. For their review, an Interdepartmental Committee on Program Balance, with appropriate subcommittees, may be necessary for critical screening and review, in the manner followed by the Requirements Committee in the War Production Board during World War II. The objective would be to plan for a balanced program for military and civilian requirements, against which resource allocations could be planned.

We realize, of course, that the problem is so great and so complex that we cannot hope, in time of peace, to do more than to design the machinery; to train a nucleus personnel within the National Security Resources Board and other agencies; and through peacetime operations produce, as the factual basis for mobilization planning decisions, balance sheets covering only key and major end-products and components, and their translation into requirements for critical raw materials, fuel, power, transportation, and manpower. Through this continuing balance sheet operation we would serve four basic mobilization planning purposes:

1st. We would provide systematic measurement of the relationship between resources and mobilization requirements and thus provide a factual foundation for advice to the President on economic readiness measures.

2nd. We would develop procedures and techniques which in the

event of war would be available for determining in detail the hundreds of resources-requirements balance sheets on which the war agencies could plan their operations.

3rd. We would develop procedures and interdepartmental administrative machinery for analyzing and integrating these balance sheets to achieve program balance.

4th. We would train a skeleton staff within NSRB and in other agencies of government which could quickly be expanded to conduct this vital operation.

It may properly be asked where the statistical records can be found for the economic mobilization planning which I have described. The general answer must be that in time of peace these must come from existing records and established channels for the statistical reporting of economic intelligence. Special mention, however, should be made of the valuable statistical materials to be found in the reports and records of the War Production Board and other war agencies. Of particular value are the records on bills of materials for translating end-product and component requirements into their raw material content. It may be added also that the Armed Services are adding to our bills-of-materials records through special provisions in procurement contracts.

Let us now consider briefly the nature of statistical requirements in time of war. Our experience in two world wars has shown clearly that effective prosecution of a major war is not possible without resort to numerous economic controls. National policy then demands that the sum total of business operations accomplish specific and well-defined objectives. To meet these objectives, government is forced into day-to-day relations with business units. Government administrators, as a result, must collect data showing a balanced factual picture of the economy. Business management requires a continuing flow of accurate and detailed information to make the kinds of decisions needed to conform to government regulations and to effect the maximum contribution of the concern's resources to the national objectives.

Statistics are needed to devise these war-time controls and to serve in their enforcement. To determine the nature and the extent of controls of manpower, for example, we need detailed statistical information on geographic distribution of employment, on the availability or non-availability of specific skills for specific industries in specific areas. Likewise, industrial controls cannot be established and effectively enforced without detailed information on needs, on usage, and on inventories. This information must be far more detailed and specific than that ordinarily available to the federal government. This information

must give us facts not only on the quantities of general types of commodities and products but on their varieties and shapes. It must give accurate indication of the physical location of inventories, their ownership, and their intended use. Similar detailed information is needed on capacities, production plans, and schedules, on storage and transportation, and on a multiplicity of other activities and needs. The collection of such information must be centrally planned. Its analysis and interpretation must be centrally coordinated.

Once initial controls have been established, detailed statistics are required as tools for the daily tasks of administering those controls. It is then that government finds itself in business on a large scale, buying and selling enormous quantities of materials and equipment. Conducting business on such a scale calls for managerial statistics which are intimately related to the accounting records from which such statistical information is derived. This type of statistics is in daily use by the management of private enterprises, but in time of peace the collection of such statistics is generally beyond the province of the federal government. In time of war, however, the federal government directs a large share of the economic activities of the nation, closely controls many of them, and must inject itself in many day-to-day transactions between sellers and buyers. Statistics required by the government to perform these functions naturally are not peace-time voluntary reports but the mandatory collections of a mass of detailed facts the reporting of which is justified by the emergency of war.

Perhaps the simplest way of indicating the magnitude of the problem of war-time control statistics is to note that the *Catalog of War Production Board Reporting and Application Forms* required nine large-sized volumes aggregating more than 3,000 pages merely to reproduce the most useful 1,200 of the 4,400 questionnaires developed and used by that agency; and an additional volume to list alphabetically the 5,000-odd commodities reported on in these 1,200 forms.

As stated in the General Introduction to that *Catalog:* "To mobilize our production resources for war, to insure that the products of mines, forests, factories, and chemical laboratories were utilized efficiently, to integrate materials and components into the greatest possible volume of finished products and to channel the distribution of finished products to the military and export agencies and to the domestic civilian population in the manner most directly related to winning the war, the WPB assembled a quantity of information of unprecedented magnitude and detail. The statistics collected included data on basic materials, semi-fabricated materials, components, subassemblies, and finished end

items. They also provided related information on the purpose for which the products were used and the kinds of ultimate consumers who used them."

It would, of course, be dangerous to assume that the statistical activities of the emergency agencies in World War II represented models of efficiency. Quite to the contrary, students of the problem can cite numerous examples of bad planning of statistical reporting and the collection of voluminous facts which sometimes could not even be tabulated fast enough to serve as guides to administrative action in a rapidly changing situation. It is, therefore, incumbent on statisticians engaged in mobilization planning to look with a critical eye on the statistical reporting of World War II; to seek earnestly to learn from that experience through selection of the good patterns and discarding of the wasteful and unnecessarily complicated patterns; and to strive for simplicity in designing the statistical controls for a possible future emergency.

THE WAR PRODUCTION BOARD'S STATISTICAL REPORTING EXPERIENCE, V AND VI

DAVID NOVICK
University of Puerto Rico
AND
GEORGE A. STEINER
University of Illinois

This is the last of a series of four articles, in six parts, concerned with the statistical reporting experience of the WPB and its predecessor agencies. In this article there is presented an analysis of a few of the outstanding tabulation problems of the WPB and its principal tabulating agent, the Bureau of the Census. In addition, the series is concluded with a statement of a few outstanding lessons which the WPB's statistical experience has revealed to be strategic in the establishment of a suitable emergency industrial statistical reporting program.

PART V

TABULATING DATA

INTRODUCTION

THE SUCCESS with which the WPB was able to acquire accurate data and to compile them quickly into meaningful summaries determined in a very real way the effectiveness of basic policy decisions and administrative control actions. The tabulation of statistics should be an elementary process. Under wartime pressures and data demands, however, there were hidden in this apparently simple process many dangers. The greater the volume of data to be tabulated and the closer it related to operations the greater these dangers became. Improper arrangement of even the most minor details in final summaries many times invalidated an entire tabulation and the details from which it was compiled. Such failures produced a lack of confidence in a series, making the figures ineffectual for the purposes designed.

Principles and procedures used to aggregate statistical series into summaries are basically the same for both great masses of data and figures in small volume. But the WPB came to learn that problems incident to massive statistical compilations, particularly when they were to be used for administrative rather than purely statistical purposes, were of real significance to its effective operation. The main thesis of this article is to study a few outstanding tabulation problems facing

the WPB in its attempt to aggregate masses of data for managerial functions, to show how they were solved, and to note the impact on WPB management.

Wartime Complication of the Tabulating Process

The wartime tabulation process was complicated by three elements: (1) pressures of time, (2) the volume and complexity of numbers demanding addition, and (3) special wartime tabulation needs which required adaptation and often important modification of peacetime mass tabulation techniques. These forces became the more difficult to control because of severe limitations of time and of large-scale tabulation facilities and experience.

Wartime operations pressed on the time cycle of statistical processing from every direction. The result was an acceleration of all steps in tabulation procedure. The greedy demands of wartime operations were never satisfied and all energies were constantly directed towards effecting a further tightening of the tabulation time schedule.

The avalanche of statistics collected by the WPB severely strained available resources for aggregating them and gave rise to a series of problems. Mechanical and personnel resources were at times strained to the breaking point.

Wartime operational needs injected new problems for both industry and government into peacetime tabulation techniques. For industry they necessitated the establishment of new records, finer detail in records, an unremitting flow of information, and quick filing of reports. The result naturally was reporting inconsistency, incompletion of forms, failure to use the specified measures and outright clerical errors of serious proportions. Large-scale tabulation methods faced the problems of quickly unifying this extraordinary mass of diverse information and aggregating it in a way often different from that specified on the basis of past experience. The penetrating eye of war sought much more detail in summaries and greater accuracy than was required under ordinary circumstances. These were not the result of unreasonable expectations of the novice; they sprang from the urgencies of immediate administrative action which probed deeply into the heart of industrial production and distribution.

Two fundamental results grew out of these considerations. First, the care born of leisure could not be devoted to the methods and devices by which basic data requests were formulated and answered. Second, time was not available to permit the careful planning of all aspects of the mass tabulation procedure. Even so, many of the new

problems were matters which could not be resolved by analysis of peacetime experience but had to be answered by trial-and-error methods. The net result was the development of one more strain on the entire administrative process which in turn multiplied and hardened still further the compulsions of time.

The Production Requirements Plan Tabulations[1]

In a very real sense, form PD-25A, the operating instrument for the PRP, embodied most of the more serious tabulation problems encountered by the WPB and its most important tabulating agent, the Bureau of the Census. For this reason it is informative to study in detail the process by which form PD-25A was tabulated, especially the nature and technique of its product coding. For this purpose we shall choose the tabulation for the fourth quarter 1942 PD-25A upon which were reported second quarter 1942 actual metal use and fourth quarter 1942 critical material requirements. Although this may be considered typical it should be mentioned that various changes in the form itself; manufacturer and WPB experience in processing; and changes in the variety of final tabulation requirements—all resulted in slightly different tabulation techniques and problems for each PD-25A quarterly tabulation.

The first step in the tabulation of PD-25A was sorting schedules to facilitate coding, review, and tabulation. The most important sorting was for size. In view of the fact that it was impossible to tabulate within the time available all PD-25A schedules received (about 35,000), smaller plant schedules were sorted out and marked for exclusion from tabulation. If a company reported less than 30 tons of total steel or 10,000 pounds of all other items together its schedule was rejected for purpose of tabulation. The reason for this treatment grew out of studies made regarding the extent of metal concentration among manufacturers and the fact that it would have been impossible to meet a tight time schedule if the tabulation included the thousands of such small cases. On the basis of PD-275 returns it was calculated that the small cases rejected and the thousands of small plants not required to file PD-25A, consumed no more than five per cent of the metal used by all plants meeting PD-25A reporting standards, as defined in priorities Regulation No. 11. The small cases which were rejected consumed less than one-half of one per cent of the total material used by all industry.

[1] For a summary of the operation of the Production Requirements Plan, see the first article in this series, Volume 43, Number 242, June 1948, p. 220.

The next step in tabulation related to product coding of the schedules to be included in the tabulation. It was in this area that the major problems arose in PD-25A tabulations and their use.

The purpose of a product code is to assemble products which are the same, similar, or closely related, into fairly homogeneous groups by assigning the same number to all items included within the same group. These numerical symbols identify the product groups throughout the process of tabulation thereby insuring final summarization for comparable products. In the process of summarization, figures representing material consumption by producers of edge tools, for example, are meaningful only if it is reasonably certain that the figures include all reports for producers of certain specified edge tools. Uniformity of concept of product groups becomes indispensable when condensed reports by commodity groups are required without listing individual items. Under these conditions the choice is a simple one: uniform product classification or completely useless tabulation.

Various government agencies, particularly the Bureau of the Census, for many years had devoted considerable attention to the proper classification of industrial data. But this work, as discussed in Article III, Part IV, was not suitable for wartime use and as a result new classifications had to be developed. The fruition of this work, it was previously pointed out, resulted in the product codes used to tabulate PD-275 and PD-25A. Under the code worked out for tabulating these forms, it was possible to classify production data with the specific product to which they related and to provide summary information which the WPB could use for product administration.

The heart of the difficult coding problem of the PD-25A tabulation lay in the desire of WPB Divisions to obtain as fine product details as possible. The basic limitation on product group expansion, however, lay in the extent to which manufacturers themselves reported detailed data by individual products. PD-25A applications were developed by a manufacturer upon the smallest operational unit which his experience led him to establish. For the most part these units reflected common experience in terms of fabricating facilities, labor problems, processes, materials used, or products made. As a result, there existed in each industry a substantial group of closely related inventory control units, even though the ultimate use of the output may have been as dissimilar as vitreous enamel kitchen utensils and vitreous enamel hospital utensils or Army and Navy mess equipment. Summarization of PD-25A data for policy and administration was limited by the extent to which current industry records could relate material use, inventory,

consumption, and requirements to detailed products. Efforts to depart very far from the reports themselves could easily invalidate the final tabulations.

Hindsight leads to the conclusion, however, that a substantial educational program explaining the needs of the PRP and PD-25A might have provided the needed inspiration for manufacturers to report finer product details. The PD-25A, covering requirements for the first quarter 1943, contained such a statement of the product detail desired. The results permit the conclusion that had such information been given industry from the time of the first PD-275 the end result might have been far finer product classes than were finally developed.

Each PRP schedule contained three product codes: the minor, the secondary, and the major product code. Each code was based on 230 product groups (later expanded to 440) and was determined by the description of the products reported in Section B of the PD-25A application, shown in the accompanying illustration. The minor code represented the finest product detail reported and made possible a tabulation of dollar shipments for each product, the summarization of which was completely related to a homogeneous class of products. The secondary product code, the most significant of all, represented the most detailed breakdown of material data in relation to product groups. It was this code, into which the vital data for WPB material distribution operations were tabulated, that set the patterns of over-all metal authorization for groups of products, and it was this code which created most of the controversy surrounding PD-25A tabulation. Its exact composition, therefore, deserves further comment.

Product shipment information was reported in Section A of PD-25A. Total detailed metal consumption, inventory, and requirements for all products produced from materials included in the smallest inventory unit of the plant were reported in Section E, Part 1, and material usage and requirements for specific products were presented in Section E, Part 2.

If a manufacturer reported in Section E, Part 2, information for each product specified in Section A, the minor and the secondary codes would have been the same. But if, as was more common, a manufacturer reported fewer products in Section E, Part 2, than in Section A, the two would be different. This happened because manufacturers reported product details in Section B which could not be related to materials in Section E, Part 2. A producer of circuit breakers, for example, might report shipments for a variety of types in Section B but lump main types in Section E, part 2, simply because materials used

Attach a catalogue or other descriptive literature, if not now on file.

DO NOT USE

Form PD-25A
5-4-48

UNITED STATES OF AMERICA
WAR PRODUCTION BOARD
WASHINGTON, D. C.

APPLICATION FOR PRIORITY ASSISTANCE UNDER PRODUCTION REQUIREMENTS PLAN

For Calendar Quarter October-December 1948

Submit six copies

Read instructions in Section A carefully before filling out Form

Form
No.
Coded
Entered
Tab.

Serial No. | Name and address of Company

Plant to Which This Report Applies | Location of Plant (Street and number, city, State)

If Entire Plant is Not Covered by This Report, State Department or Operating Unit Covered

SECTION B, *Part 1.*—**Actual Shipments:** Show below the number of units and dollar volume of shipments (including transfers to other divisions, departments, or plants) of each of the classes of products shown in column (1). Please read the instructions in Section A for guidance in grouping the products you manufacture. Entries must appear in columns (3)–(7) and (9)–(10). Where the answer is None enter the word "None."

DO NOT USE	Description of Classes of Product (1)	DO NOT USE (2)	Actual Shipments During Calendar Quarter April–June 1948								
			Total Dollar Value (Unit Costs) Analyzed by Preference Ratings								
			Total (3)	Allocated or specifically directed (4)	AA (5)	A-1-a (6)	A-1-b through A-1-k (7)	DO NOT USE (8)	A-2 through A-9 (9)	Other and unrated (10)	DO NOT USE (11)
	Total										

SECTION B, *Part 2.*—**Estimated Shipments.** Show below in column (3) total estimated shipments for the calendar quarter October–December 1948. Enter in columns (4)–(10) only *rated* orders now on the books scheduled for shipment during the quarter. Enter in column (11) the balance of estimated shipments for which no rated orders are on hand. See instructions, Section A. Entries must appear in columns (3)–(11). Where the answer is None enter the word "None."

DO NOT USE	Description of Classes of Product (1)	DO NOT USE (2)	Estimated Shipments During Calendar Quarter October–December 1948								
			Total Dollar Value (Unit Costs) Analyzed by Preference Ratings								
			Total (3)	Allocated and AAA (4)	AA-1 (5)	AA-2 (6)	AA-3 (7)	AA-4 (8)	A-1-a through A-1-k (9)	A-2 through A-9 (10)	Other and unrated (11)
	Total										

SECTION B, *Part 3.*—**Total Unfilled Rated Orders:** Show here the total dollar value of unfilled *rated* orders as of the most recent date available.

Total, $_____ as of _____ (Give date)

SECTION C, SECTION D } Information called for by these sections in previous PD-25A forms is not required.

PD-25A
6-4-42

(Name and location of plant) (Department or operating unit)

SECTION E, *Part 1.*—Report all materials shown on Materials List No. 1, Revised, used in the manufacture of products listed in Section B, or for maintenance and repair. When materials are used in the forms shown on the portion of the Materials List printed below on pages 2, 3, and 4, enter quantities opposite the appropriate items. Quantities must be reported in the units indicated in column (2). Enter other items on the Materials List on black page following the printed portion of the list, using the units specified in the Materials List. Even though you do not keep your records in these units, please convert into the units specified. Report quantities in columns (4) through (8), except for chemicals (see Section A, par. 32).

Item No.	Description of Materials (1)	Repeat Item No. (2)	Unit of Measure (3)	Statement for Quarter April-June 1942: Received during quarter (4)	Statement for Quarter April-June 1942: Put into production (used during quarter) (5)	Statement for Quarter April-June 1942: Inventory on June 30, 1942 (6)	Estimated Inventory on September 30, 1942 (Optional) (7)	Statement for Quarter October-December 1942: To be put into production (Anticipated requirements) (8)	Statement for Quarter October-December 1942: DO NOT USE (9)	Statement for Quarter October-December 1942: DO NOT USE (10)
	IRON AND STEEL			Quantity	Quantity	Quantity	Quantity	Quantity		
	Iron:									
100	Pig iron		Short tons							
	Castings (before machining):									
101	Gray iron		Short tons							
102	Malleable		Short tons							
103	Pipe		Short tons							
104	Bars and rods, wrought		Short tons							
105	Pipe and tubes, wrought		Short tons							
107	Plates and all other, wrought		Short tons							
108	Cast iron scrap—inventory figures only		Short tons							
109	Steel scrap—inventory figures only		Short tons							
	Carbon Steel:									
110	Ingots, blooms, billets (except shell steel), slabs, sheet bars, and tin bars		Short tons							
111	Billets, shell steel		Short tons							
112	Tube rounds		Short tons							
113	Structural shapes and piling		Short tons							
114	Plates, including skelp in plate sizes		Short tons							
117	Skelp (except in plate sizes)		Short tons							
118	Rails, over 60 pounds		Short tons							
119	Rails, all other		Short tons							
120	Tie plates and track accessories		Short tons							
	[illegible] bars, except shell steel and con-		Short tons							
			Short tons							
			Short tons							

PD-25A
6-4-42

(Name and location of plant) (Department or operating unit)

SECTION E, *Part 2.*—This section must be filled out if more than one class of products is reported in Section B. See Instructions, Section A. Report all materials for which entries appear in columns (5) or (8) of Section E, Part 1. Report materials in the units specified in column (3) of Section E, Part 1. Dollar values are not to be reported in this section.

Item No.	Description of Materials (1)	Repeat Item No. (2)	Unit of Measure (3)	Consumption and Requirements for Separate Classes of Products: Products: Describe each class separately A			B		
				Put into production April-June 1942 (4)	To be put into production October-December 1942 (5)	DO NOT USE (6)	Put into production April-June 1942 (7)	To be put into production October-December 1942 (8)	DO NOT USE (9)

for detailed types came from the same inventory bin. The problem may be explained more fully in this way. Suppose the Ajax Corporation reported in Section B, as follows:

Products	Dollar Shipments
Jigs	$10,000
Dies	10,000
Other machine tool accessories and parts	10,000
Electric conduit metal clamps	2,000

Each of the above items would be given a minor code. The first three, being items of the same genus, would carry one secondary code, and the fourth item another secondary code. Now, suppose Ajax either did not provide a Section E, Part 2 for any of these items, or combined the fourth item with one of the first three. Either of these alternatives would have raised a serious coding problem. The following details should have been reported for clean-cut tabulation: (a) use and requirements for dies, jigs, and other machine tool accessories and parts, and (b) use and requirements for electric conduit metal clamps.

In a situation such as this, several alternatives were open. The metal data for all products, including that for electric conduit clamps, could have been thrown into a product group the composition of which was solely devoted to information about machine tool accessories, since it was clear that the major products of the company fell into this category. For a small plant this would have had little importance. If the Ajax Corporation, however, had been larger than it is shown to be, a decision to do this would have seriously affected the accuracy of the over-all tabulations. It would have impaired the accuracy not only of the product group designed for machine tool accessories (which would have been loaded with other irrelevant data) but also of the product group concerned with electric conduit metal clamps (which would not have reflected the true volume of materials used and required for this product).

In view of this situation, of course, every effort had to be bent toward segregating information for both groups. This sort of problem was solved on the basis of information submitted by the plant on preceding questionnaires; from the knowledge of the Bureau of Census technical experts; from WPB Division personnel having cognizance over the plant; or from direct correspondence with the company. Frequently,

the schedule itself furnished information by which good estimates of the true situation were calculated.

Finally, a major code was given to each schedule to indicate the major product class of the plant or principal product as determined by dollar shipments. This code was necessary to tabulate receipts, use and inventory information for which no secondary product group tabulation was possible. The form did not provide for such a distinction of inventories. This decision in formulating the questionnaire, of course, was dictated by industrial practices and the tremendous difficulties which efforts to do otherwise would have imposed upon industry.

Other codes were placed upon each PD-25A schedule, such as a geographic code to permit tabulation by States and WPB Regions; a WPB Industry Division code, to show which section of the WPB would process the form and to provide it with a tabulation of concern to it alone; and various other codes designed for technical purposes, such as a completion code, a size code, etc. All of this coding procedure was dictated by the complexity of the final tabulations required.

Review and editing of PD-25A constituted a large and important operation. The principal functions of editing were to review information reported, to detect errors and correct them, and to prepare the schedule for machine tabulation. The report was surveyed to make sure that it was in usable form, was accurate and consistent with instructions, and would not disturb the over-all accuracy of the final tabulations. It is clear that tabulation of unedited or unreviewed schedules would have resulted in completely meaningless tabulations.

There were many instances where data reported on the PD-25A needed adjustment, but only a few of the less technical points need be treated here. Shipments were sometimes reported only in units rather than units and dollars; sometimes shipments for a future quarter were reported as smaller than those for a past quarter yet anticipated metal requirements for the advanced quarter were substantially beyond consumption in the past quarter; shipments reported on schedules of subsidiaries sometimes covered the entire company of which it was a part, while metal figures related only to the subsidiary; and so on.

Greater problems were encountered in connection with the metal figures. Sometimes manufacturers would report tons where pounds were required, and vice versa; frequently forgings and castings were reported in units, pipe in feet, and wire in rolls. Clerical errors on the part of plant personnel resulted in omission of important requirements and placing information on the wrong line or in the wrong column; and often odd material sizes and shapes were written in the preprinted stub.

Other typical and more difficult errors to correct were as follows: subcontractors, contrary to instructions, often reported material that they did not own but processed for a prime contractor; in some instances a manufacturer making basic metal shapes and forms specified on the preprinted stub, forgings for example, reported metal consumption and requirements for his finished basic product; manufacturers often reported one thing on PD-25A and in attached correspondence described conditions which necessitated eliminating the figures; and many times material reported in Section E did not appear in Section E, Part 2, when production clearly indicated the material had to be used to make the product specified. Obviously these errors had to be corrected for tabulation machines added integers only and produced figures no better than those which went into them.

In all tabulations it is important to know the exact part of the universe which the questionnaire is presumed to cover. Two problems in this connection presented themselves in tabulating PD-25A applications. First, it was necessary under the operation of the Production Requirements Plan to consider total requirements for critical materials and to compare them with total estimated supply. Since the PD-25A applications themselves were not used as application forms in certain non-PRP or "exempt" areas the WPB had to make arrangements to acquire "master PD-25A" schedules covering material use, requirements, and inventory for consumers in these areas. This was a WPB rather than a Bureau of the Census Problem. Second, in the absence of a long experience with the PD-25A and its predecessor PD-275, and because of changing war production and industry experience in filing this type of report, it was exceedingly difficult for the Bureau of the Census to know when sufficient reports were received to assure reasonably complete coverage in particular industries or for all industries. Tight tabulation time schedules necessitated stopping the tabulation procedure somewhere so that schedules could be added and presented to the WPB for its administration. But tabulations incomplete in totality or in important particulars could not be adapted readily to administrative needs.

Obviously one of the first important steps in assuring completeness was effective follow-up procedures for stimulating quick industry response. But beyond that the Bureau of the Census and the WPB had to develop measures of completeness for the entire tabulation and important product groups thereof. The most important and practical measure was a list of important plants that had to be included in the tabulations to assure completeness. The high degree of material con-

centration among a relatively few large plants facilitated the determination of an estimate of the extent of coverage of the total tabulation merely by checking the important plants included. Within narrow product groups this technique, although applicable, was not so practical. And no matter how carefully incoming schedules were matched against lists of important consuming plants, estimates of coverage in the aggregate or for particular products were subject to a wide margin of error and only the preliminary tabulations could reveal accurately the extent of completeness. This problem, due essentially to lack of historical series for this type of data, created an unnecessary risk in both tabulation and WPB operations.

Following the above processes, schedules were forwarded to the machine tabulation unit where punch cards were prepared. One card was punched for each line of Section B, one card for each item in Section E, Part 1, and one card for each item in Section E, Part 2. In addition, there was punched on each card the major product code, the secondary product code, the minor product code, the WPB Division to which the schedule was assigned, a size code, a completeness code, and a serial number for the schedule. Altogether approximately 500,000 cards were punched for the fourth quarter, 1942, PD-25A tabulations. In the next succeeding quarter the volume of cards almost reached 800,000. Preliminary tabulations were prepared from the cards and subjected to critical review. In this process, the figures were given additional careful scrutiny to detect errors appearing when the data for one homogenous group were listed together. Literally miles of tabulating machine listings were produced from these cards.

The foregoing account leaves little doubt that tabulating PD-25A schedules not only was a monumental undertaking but was made in the face of serious technical problems. The time schedule for tabulation frames the problems of tabulation *multum in parvo*. For the fourth quarter 1942 the time schedule was as follows: receipt of schedules, 25 October–10 November; first machine listing, 11 November; machine listing for the WPB Requirements Committee, 16 November–23 November. Here was what amounted to a census of the metal-working trades, completed in less than one month!

Despite the immense effort to adapt tabulation procedures to fit administrative methodology, the PD-25A tabulations and those of its predecessor PD-275, raised loud criticism in the WPB. Censure sprung from many factors, perhaps the most serious of which was the coding system used by the Bureau of the Census. The system of coding was necessarily rooted in a classification system which reflected existing

industrial organization. The development of the WPB Division structure, on the other hand, was in terms of product assignments which in many instances did not reflect the real organization of American industry.

These divergent factors created a fundamental conflict between the tabulated summaries and the WPB's ability to use them effectively. The net effect was a basic policy decision on the basis of tabulated summaries which could not be carried out fully by the WPB Divisions. The Census, for example, was obliged to tabulate insulated wire and cable into one category partly to obtain a clear-cut homogeneous group of data and partly to permit tabulation of individual plant schedules reporting production of insulated wire and cable, without reporting detailed data for all types and sizes. But in the WPB various types of insulated wire and cable were distributed and managed by different divisions with conflicts in the jurisdiction and responsibilities of each. The Copper Division was assigned various types of copper cable; the Communications Division was concerned with coaxial, lead-covered, telephone, and submarine cable; the Shipbuilding Division was interested in armored or degaussing cable; and the Building Materials Division processed requests for armored cable, electric cable, and insulated cable. Other products, such as electric motors, valves, and condensers were each under the cognizance of many WPB Divisions. Some important products such as storage batteries; turbines and water wheels; bolts, nuts, washers and rivets; and miscellaneous stamped and pressed metal products, were completely unassigned to any WPB Division. No juggling of tabulations could adjust to this illogical assignment of products within the WPB itself.

There were two further problems. The WPB Divisions, partly to fit tabulations into their organization and partly to fit them into their preconceived notions of the type of data summaries needed to provide efficient administration, felt that tabulated summaries should be made in much greater detail than was possible with the 230 product groups used in tabulation. A simple illustration of the problem relates to one product group established for "plumbing fixture fittings and trim." Into this group the Census tabulated products including victory trim which contained a small amount of copper alloy; brass trim which was 100 per cent copper alloy; plumbing specialities which might have been either; pipe hangers, supports, and rests, which were largely carbon steel; and other fixtures of varying types of metal content. The WPB problem centered about the policy determination of allocations of this group and Division and industry compliance with basic policy. With

such a farrago it was difficult for the Requirements Committee to know how much carbon steel included in the allocations for this group would be consumed in victory trim as against pipe hangers. Likewise the responsible Division had no method for determining, without special inquiry, what proportion of the metals allocated to this product code would be in victory trim (which was available to the public), or brass trim (which was available only to the military services). The results led to WPB pressures for finer and finer product detail. The Census problem in accomplishing finer product detail, of course, centered about limitations inherent in manufacturer reports. In the face of limitations imposed by the reports themselves the effort to produce more detailed summaries intensified the entire tabulation problem and subjected it more and more to error. In later quarters the product groups were refined but never to the complete satisfaction of the WPB Divisions.

A second type of problem pertained to the Census policy of attempting to make product group data homogeneous. But this policy, particularly in the first quarter of 1943 tabulation, met head on with increasing difficulties in the WPB in relating the chain of production to final products. The tabulating agent was asked, therefore, to code products into their end use categories whenever possible. This was done for the first time in the first quarter of 1943 PD-25A tabulation. Thus, for example, brass valves for ships, where identifiable as such, were set up in a separate product group under ships; valves for track-laying tractors were tabulated with track-laying tractors; low pressure hydraulic valves for aircraft went into aircraft; boiler feed regulator valves for locomotives were tabulated in locomotives, and so on. Once again, the principal limitation was in the plant reports themselves. Since a complete end-use identification was not possible for all components the resulting final summaries were rather far from homogeneous and did not completely reflect incorporation of component product data into product groups established for their respective end items.

All these difficulties were accentuated because of simple errors in the schedules themselves. Since the coding experts in the Bureau of the Census were few in number and since time pressures were so great, much of the coding had to be performed by clerks. Although errors were for the most part relatively minor, they assumed tremendous proportions in WPB debate. Natural unfamiliarity with a wide variety of trade names, many of which were introduced with war production, led to miscoding. Some homographic words were miscoded, such as concertina barbed wire under musical instruments and fire control instru-

ments under fire-fighting equipment. Such errors were magnified out of all proportion to their significance by WPB personnel and eventually led to rather widespread dissatisfaction with PD-25A and PRP.

Decentralized Tabulation of Operating Data under the PRP and the CMP

The PRP experience lent a deeper and stronger support to those WPB material distributive systems which followed it than most people are willing to admit. The CMP, for example, would not have been possible without the experience gained under the PRP. It is in the realm of tabulation that one of the supports for this conclusion is clearly revealed.

The PRP, as a mechanism for material distribution, brought to light many important "bugs," the spotting of which gradually filtered up through the various managerial levels of the WPB and on the way crystallized into rather clear-cut issues, which could be resolved only by top-side policy decisions. The more important questions related to the precise material shapes and forms to be controlled by a master distributive machine, the exact organization of summary data for policy decisions and administration, and the methods by which data were to be accumulated on both the requirements and authorization side of the management chain of action. These issues were resolved in the promulgation of the Controlled Materials Plan.

Conclusions of the planners and administrators of basic policy coincided in the fourth quarter of 1942 concerning the need for and advantages of decentralized tabulation of authorizations made by the WPB Divisions on PD-25A and other material allotment instruments. It was recognized that the WPB could no more administer its program for metal distribution without some accounting of action than a private business enterprise could carry on its day-to-day operations without some accounting for its cash withdrawals. In addition, it was recognized that decentralization of the authorization tabulations within each WPB Division offered many administrative advantages over centralized accounting control tabulations.

These concepts developed into a number of basic operating tenets. The following were of considerable importance: (1) those responsible for administration should also be made responsible for the figures which formed the basis of control action; (2) decentralized control records should be available for uses other than control accounting; (3) specialized problems relating to particular industries, plants, and products could be best ironed out on a local basis; (4) disagreements over methods of tabulation and accuracy of results disappeared under

decentrally posted but centrally directed tabulation; and (5) requirements data and authorization actions could be readily related through all levels of administration for each detailed segment of industry when data were tabulated decentrally. Undoubtedly, centralized tabulation had certain advantages over decentralized tabulation but administratively the latter technique was far superior to the former.

To test the statistical validity of these tenets both the Bureau of the Census and the WPB Divisions tabulated fourth quarter 1942 PD-25A returns. Substantial and ridiculously careless errors were detected in the WPB Division tabulations when the two series of data were matched. Nevertheless the experience was well worth the effort and firmly established both the utility and feasibility of WPB Division decentralized tabulation of data relating to the operation of a comprehensive metal distribution program such as the PRP.

In the meantime, plans were under way for the introduction of the Controlled Materials Plan in the second quarter of 1943. Tabulation techniques contemplated in the operation of the CMP fortunately were guided by the previous history of the PRP and other tabulation experiences. First, a firm decision was made concerning the metal shapes and forms to be included on the stub of the CMP application forms. Second, the WPB firmly grappled with the problem of product codes into which Class "B" products were to be incorporated. And they were fitted into the WPB's operations. This stilled much of the WPB criticism of Bureau of the Census product codes used in the tabulation of PD-25A. Third, the problem of proper editing of forms before tabulation was recognized and acknowledged. Finally, the WPB Divisions themselves were to make both the original requirements tabulations and the tabulation of actions taken on the basis of the WPB Requirements Committee decisions.

The problem of determining product groupings was divided into two parts, reflecting the combination of vertical and horizontal material distribution techniques upon which the Plan was based. On the vertical side, the Claimant Agencies determined their own groupings in terms of final products, or programs. The WPB problem with respect to the Class "B" list was comparable with that encountered under the PRP but of less formidable dimensions. Here classification had to be in terms of fabricated products entering into end products. To grapple with this problem the WPB established a Product Assignments Committee which attempted to reconcile the differences existing among the various Divisions, to establish firmly individual Division responsibility over an individual product, and to place responsibility for groups of

related products in one Division. Experience under the PRP went a long way toward defining properly the individual items to be included in a particular Class "B" code and in insuring that items calculated to be related were really homogeneous and reportable by industry in the terms specified. But despite this experience, there were some difficulties with the listing. There was some overlapping of responsibility among the Divisions, some product groups were too refined for good industry reporting and effective WPB processing, some product groups were too narrow and not representative of a homogeneous group of products, some products were completely overlooked, and some product groups were far too inclusive. Altogether, however, the difficulties raised in the original listing were not of substantial proportions and operations over a few quarters rather well ironed out the problems.

The classification scheme established in February 1943 for Class "B" products embraced 484 product groups as compared with 441 used under the PRP in the first quarter of 1943. This represented a substantial expansion because it will be remembered that the Class "B" listing was confined to presumably shelf items and did not include the industrial universe as did the PRP. A great volume of end products such as guns, tanks, ships, etc., were considered Class "A" products and taken care of in the programs established by Claimant Agencies. It is interesting to note that some PRP product groups were combined to form one CMP Class "B" group. Such instances, however, were few.

The history of product group classification up to this point was one of general expansion, modification, consolidation, and adjustment. This process was so widespread that any attempt to determine summary data covering the war period from the various instruments used by the WPB was greatly complicated by variations in product groupings, not to mention coverage, type of data obtained, and overlapping time periods. Thus, tabulation of combinations of reports was virtually impossible. The introduction of the CMP classification systems made data obtained from the series of comprehensive reports preceding it virtually useless in the operation of the CMP.

Preceding experience with authorization instruments demonstrated beyond doubt that to obtain good and accurate tabulations of requirements and to furnish the basis for accurate accounting it was necessary to subject incoming forms to some critical review. During the second quarter of 1943, the first quarter of the CMP operations, the Bureau of the Census tabulated CMP-4B requirements for the WPB. But in the third quarter of 1943, when the WPB Divisions first undertook that responsibility, the manual of instructions to the Divisions contained a

careful analysis of the types of adjustment necessary to provide uniformity in tabulation and program determinations. Recognition of and planning for inaccurate reports spelled the difference between usable and unusable WPB summaries.

Fundamentally, the problem of the WPB Division tabulation of controlled material requirements and authorizations against those requirements was simple. It was merely an additive function. The basic reporting requirements for specific metals and metals shapes and forms were well established; the problem of specifying end-use on the forms had been resolved by means of the extension of program symbols; the product classifications into which the data were to be summarized and assignments of the product groups to specific WPB Divisions had been accomplished with success; and finally, industry had ample opportunity to adjust its records to provide the sort of data needed on various application instruments. There was little else to do but add.

The WPB Industry Divisions were not prepared to tabulate material requirements on CMP-4B applications submitted for the second quarter of 1943, and the Bureau of the Census was asked to make summaries for the WPB. To facilitate operations the WPB received the applications, coded and edited them, and transmitted them to the Bureau of the Census for aggregation. Because of the preceding experience with the PRP, however, the WPB Divisions were quite able to maintain authorization accounts for the second quarter of 1943, and all subsequent quarters. In the third quarter of 1943, both the WPB Industry Divisions and the Bureau of the Census tabulated incoming applications, the dual tabulation constituting insurance against WPB Industry Division failure. As it turned out, a series of audits among the WPB Industry Divisions showed that tabulation was proceeding accurately and satisfactorily. As a result the Bureau of the Census tabulation was not completed. In the fourth quarter of 1943, WPB Industry Divisions assumed all CMP requirements and authorization tabulations with which they were concerned, except those pertaining to mill reports of shipments. These were tabulated by the Bureau of the Census.

Under the operation of the Controlled Materials Plan during the fourth quarter of 1943, which procedure continued throughout the war period with but minor modifications, the WPB Industry Divisions were responsible for a number of tabulations. The most important related to a summarization of material requirements reported on incoming applications, and accounts showing authorization made against them. The Industry Divisions tabulated data on incoming CMP-4B

FORM GA-298
FORMERLY PD-37
(7-28-43)

UNITED STATES OF AMERICA
WAR PRODUCTION BOARD

SUMMARY OF CMP-4B APPLICATIONS FOR ALLOTMENT OF CONTROLLED MATERIALS FOR 4TH QUARTER 1943

DIVISION

PRODUCT GROUP

CODE | TOTAL NUMBER OF REPORTS SUBMITTED

REVIEWED AND APPROVED BY

CMP OFFICER

EXTENSION

DATE

DESCRIPTION

REPORT AS OF

1943

SECTION C - **PURCHASE SCHEDULE**

CONTROLLED MATERIALS	CMP CODE	UNIT OF MEASURE	TO BE PUT INTO PRODUCTION 4TH QTR. 1943	REQUIREMENTS 4TH QUARTER 1943
STEEL				
CARBON STEEL *(Incl. wrought iron)*	2001-2061	SHORT TONS		
ALLOY STEEL	2501-2561	SHORT TONS		
COPPER AND COPPER-BASE ALLOYS				
CU-BASE ALLOY SHEET AND STRIP	3011	LBS. OF ALLOY		
CU-BASE ALLOY ROD, BAR AND WIRE	3021	LBS. OF ALLOY		
CU-BASE ALLOY TUBING AND PIPE	3041	LBS. OF ALLOY		
BRASS MILL COPPER PRODUCTS	3051, 3061, 3071	LBS.		
WIRE MILL COPPER PRODUCTS	3131	LBS.		
FOUNDRY PRODUCTS	3201, 3211	LBS.		
ALUMINUM				
ROD, BAR, WIRE AND CABLE	4021-4121, 4151	LBS.		
RIVETS	4122	LBS.		
CASTINGS	4202-4218	LBS.		
SHAPES, ROLLED OR EXTRUDED	4251, 4301, 4311	LBS.		
SHEET, STRIP, PLATE AND FOIL	4351, 4361, 4601	LBS.		
TUBING	4401, 4411	LBS.		

SECTION A - **VALUE OF SHIPMENTS ANALYZED BY PREFERENCE RATINGS AND CLAIMANT AGENCIES**

ITEM	CALENDAR QUARTER							
	SHIPMENTS APRIL - JUNE 1943		ESTIMATED SHIPMENTS JULY - SEPT. 1943		ESTIMATED SHIPMENTS OCT. - DEC. 1943		RATED ORDERS SCHED. FOR SHIPMENT OCT. - DEC. 1943	
	($000'S)	(%)	($000'S)	(%)	($000'S)	(%)	($000'S)	(%)
TOTAL		100.0		100.0		100.0		100.0
PREFERENCE RATING								
AAA-AA1								
AA2								
AA2-X								
AA3								
AA4								
AA5								
ALL OTHER								
CLAIMANT AGENCY								
ARMY (Ow)								
NAVY (N)								
A.B.C.O. (C)								
MARITIME (M)								
O.E.W. (E)								
O.L.L.A. (L)								
CANADA (D)								
A.M.A. (M)								
O.D.T. (T)								
O.C.D. (S)								
O.P.D. (R)								
O.W.U. (U)								
F.A.W. (P)								
[illegible] (A)								
W.P.B.								
B PROD.								
OTHER AND UNIDENTIFIED								

*CLAIMANT AGENCY NOT TABULATED SEPARATELY - MAY BE INCLUDED WITH OTHER AND UNIDENTIFIED.

B-4879

applications for which they were responsible and summarized them for presentation to the General Statistics Staff of the WPB on form GA-298, shown in the accompanying illustration. These summaries were then duplicated, aggregated, and presented to the WPB's Vice-Chairman for Operations.

Some relatively minor complications in this simple additive process grew out of time pressures, the need for including large late respondents, and complicated summaries required from the Industry Divisions. There was a general tendency to increase the number and complexity of the types of information summarized by the WPB Industry Divisions. Tabulations for the third quarter of 1944, for example, were presented on forms GA-689 (Requirements) and GA-689 (Explanation). The format of these two forms presents eloquent testimony to the increasing diversity and complexity of the tabular requirements. But as the CMP operations became smoother the need for such complicated presentations vanished and in 1945 the tendency towards simpler reports again prevailed. The third quarter of 1945 CMP-4B requirements tabulations (made in June 1945), for example, reverted back to form GA-289. Although these were the basic summary presentations the responsibility still existed for the Divisions to provide a variety of other compilations.

In the operation of the CMP, as in the PRP, time schedules were always tight. In the fourth quarter of 1943, completed tabulations of almost 33,000 CMP-4B applications were made 55 days following the final date for filing the applications. Subsequent quarterly operations shaved this time cycle considerably.

WPB Field Tabulations

The real impetus to decentralized tabulations in WPB Field Offices originated with the broad decentralization policy of the WPB in 1943. Transfer of more and more operating functions to field offices at that time would have left a gap, had it not been filled, in the flow of significant information to officials responsible for the establishment of these policies. With expansion of field office actions, therefore, it became essential that they be charged with the responsibility for maintaining uniform reports on actions, consistent with and related to comparable records maintained in the WPB Industry Divisions.

Experience with decentralized tabulation, both in the WPB and its field offices, was satisfactory enough to set it apart as being feasible and practicable. Indeed, it may be concluded that sound wartime administration is better achieved when tabulation of data closely

related to administration is made decentrally by those who are responsible for the use of the data in management. This conclusion has great importance for mobilization plans of the future which must utilize decentralized administrative methodology.

PART VI

LESSONS OF EXPERIENCE

The task of the War Production Board and its predecessor agencies can be stated simply. It was to manage the industrial economy of the United States in such a way as to produce the maximum war and war-supporting goods and services with the minimum dispatch.

This was predominantly an administrative job. To an extent unprecedented in our history the national economy was guided by government between 1942 and 1945. During the war the construction of every new industrial plant was authorized by government. Not a pound of steel or copper or aluminum could legally be fabricated and used without government approval. In all industrial production and distribution the decisions concerning what should be produced, who should produce it, and to whom it could be shipped were directed by government.

Facts and proficient factual collection methods were indispensable to the intelligent and efficient performance of governmental management. Actions taken in industry were recorded on millions, yes billions, of pieces of paper. It became necessary for the WPB to aggregate information from these documents to provide the means for making informed policy decisions, to provide the instruments for administering the thousands of actions intertwined in a single policy decision, and to develop records of the actions taken on the basis of policy decisions.

Looking back on the beginnings in 1940 and 1941, and even the later developments of 1942 and 1943, two facts stand out in connection with the emergency data-collection system. First, many of those responsible for industrial mobilization and continuous control of the economy did not have a clear concept of the vital place which systematized factual reporting occupied in the emergency management function of government. Second, methodological discipline required to systematize the collection and use of facts needed for emergency management was not developed and ready for use. The result was the parallel evolution of a recognition of the need for factual accumulation to support management and the development of suitable techniques to get the necessary facts. The process of meshing the two was slow and costly.

Two classes of lessons emerge for those who in the future must deal

with the problems of emergency organization and direction of the nation's economy. The first group relates to the role played by an industrial statistical reporting system in emergency management. The second group relates to the technical methodology, and the principles which should be used in its application, in developing and presenting the facts needed for formulating emergency policy and administering that policy.

Before presenting these lessons it is worthwhile to draw attention to the applicability of World War II reporting experience to emergency problems of the future. It is our contention that no matter what the circumstances may be in a future emergency, wise management is impossible without facts; that it is possible to develop a system for collecting and using facts which is adaptable to government management of industry in all potential emergencies; and that the experience of World War II establishes the outlines and basic character of that system.

The Role of an Industrial Reporting System in Emergency Management

In the desperate search for means with which to supply the materials of war, we learned that effective decisions could be made only on a basis of knowledge. In the development of early control policies and techniques of administration it was assumed that statements of policy, supplemented by a few administrative reports, would provide the framework upon which industry could and would automatically mobilize for war production. It soon became apparent that the pull of the regular customer always diverted a substantial part of industrial output from war production unless statements of control policy were based upon understandable, complete, fool-proof facts and administrative techniques. Mere statements of policy, we learned, would not provide the guns, planes and tanks when manufacturers who supplied the goods of war were at the same time engaged in making automobiles, refrigerators, cosmetics and other items which offered them profitable use of materials, facilities and labor. We also learned that control over industry, once it converted to war production, could not be achieved by broad policy statements unsupported by concrete reporting to and from government. No firm control was possible without a method for collecting and using facts as instruments of both policy and administration.

The experience of the WPB clearly reveals the character of the industrial factual collection system which emergency management requires. Stated simply, it is necessary to have a complete and detailed

knowledge of the nation's resources and the demands which war will make upon them. This knowledge must be available on a comprehensive scale, reveal the most minute details, be in common terms by means of which the entire range of facts can be related to one another and to aggregates, and cover specific time frequencies. It means developing a national resources-requirements budget. But the factual collection system must accomplish more than that. Not only must it furnish the information essential for policy decisions but it must also provide the means through which mandates for specific action are transmitted to industry in specific terms. At every stage of responsibility this flow of authority must provide a basis for precise and definite accountability and reports of administrative action. These are the fundamentals of an adequate industrial reporting system suitable for, and flexible enough to support, emergency management.

Since the underlying problems of industrial mobilization are largely those of translating available men, materials, and machines into the greatest possible quantity of goods required for military operations and war-supporting activities, there must be a complete and penetrating knowledge of all resources. Information about resources requires a reporting system which will provide in integrated detail the actual and potential production of mines, forests, farms, smelters, refineries, factories, power plants, railroads, public utilities, and other productive facilities. The reporting system must provide knowledge of the smallest as well as the largest economic unit. The system should cover facilities for producing various types of materials, the actual level of operations, unfilled orders, potential capacity, labor requirements, material and component requirements, and so on. The entire fabric of industrial resources and their use must be revealed clearly in totality and in detail. The details must tie together in a factual picture, given in common units of measure and in definite time periods.

Against this knowledge of resources must be placed a statement of the demand for them in such terms that the two may be related. The demand for resources must be built upon a detailed, realistic and balanced statement of military, essential civilian war-supporting, and export program requirements. This statement must be in terms of specific items needed directly by the armed forces and civilians; the materials, components and facilities required for their production; and the time periods in which both final products and the intermediate products needed for their production are scheduled. In short, total requirements must be known and classified into procurement progams and production details. This factual statement must be realistic, ac-

curate, and in common units of measure which can be related to resources.

To determine the extent to which certain types of war activities could be supported, the WPB found that it required knowledge of the demands of a specific activity under consideration and information on all of the related supporting and dependent activities. Thus, in appraising the size of an ordnance program for Oerlikon guns, it was necessary to determine not only the quantities of nickel-bearing steel that would be used in the shafts of the guns but also the quantities of nickel-bearing steel that were required for the bearings of the gun, the shells that would be fired from the guns, the axles of the railroad cars that would haul the finished guns and the material used in their production. To appraise the one requirement for nickel-bearing steel all other requirements had to be evaluated, from the nickel-bearing steel used in the finished guns to the nickel-bearing steel required to make replacement bearings for a machine which mixed malted milks at a canteen in a war plant.

This sort of informational need grew out of the problems and circumstances of war. When the economic machine was straining at optimum capacity, when shortages appeared in many areas, wise direction of the machine was possible only when government on the basis of facts could choose among several demands focused at one point in the production process. If the supply of nickel-bearing steel was less than the total demand for it how else could an intelligent decision be made to double the quantity of steel entering into Oerlikon gun parts and reduce the amount of steel used as replacement parts for malted milk machines? Effective decisions required a factual review of total supply of and all demands for the steel.

It should not be assumed that once the size of the expanded war program was projected and the economic system was approaching capacity operations, the need for this interrelated information became readily apparent. On the contrary, realization of the need for facts and the magnitude of the problem of getting such facts was gained only after long and costly experience.

In the early defense period we tried to maintain all peacetime activities and at the same time satisfy the demands of war. Only when we realized how perilous was our existence and how great was our need for weapons did we have the courage to stop the production of a few unessential items. Later, in mid-1942, we belatedly recognized the fact that actual war output was too small and that we could not in fact have both guns and butter. Attitudes were rapidly switched by the

winter of 1942 and the military services were demanding programs which, if attempted, would not only have precluded butter but would have so strained the economy as to almost stop production of guns too.

In all of these actions we were driven continuously between desire and reality. The cord that joined the two was at all times knowledge. But the precise knowledge required did not exist when most needed. It was unfortunate that in the beginning we could no more foresee our needs for information than we could foresee our needs for armament. We tried to satisfy everyone in the early defense period by a free apportionment of materials and facilities and by the avoidance of a large-scale reporting burden on industry.

Ultimately we learned that we had to have facts and developed the outlines of the industrial reporting system required for emergencies. We learned that the system should produce a complete budget of all vital resources and all demands to be made upon them. The data must be complete and precise, from the supplies of basic resources to the procurement programs for final military, export, and essential-civilian products. With the data at hand it must be possible to adjust and balance programs among themselves and in the aggregate and to relate them to resources in such a way that objectives established are feasible of attainment. The data must show the time sequences in which various quantities of limited raw materials are needed at all levels of fabrication from basic material processes to final products. Throughout, the data must tie horizontally and vertically, must be expressed in uniform terms and must be related to uniform periods of time. Beyond all this, the methodology of collecting these data must be such that accurate results can be acquired in a matter of days. The more serious the emergency the more urgent becomes the demand for this sort of integrated reporting.

The more integrated and comprehensive the reporting structure the wiser will be basic policy and the more effective will be the administration of that policy. With an integrated reporting system it is possible to determine whether specific policy objectives or slightly modified ones can be achieved. It is possible to translate policy decisions into actions to be taken by large numbers of people at different administrative echelons. It is possible to convey in concrete terms to every person required to act both the specific authority and responsibility which each has and the precise limits imposed on the exercise of authority. An integrated statistical system should provide the methods and procedures by means of which information essential for policy decisions can be obtained, by which mandates for specific sections can

be transmitted through administrative levels, and by which precise accountability for the administration of each policy can be achieved.

It must be realized that an integrated industrial reporting system is not the sole basis for efficient mobilization and industrial control by government in emergencies. Time and again the WPB found that its internal organization had to be geared to the methods of the reporting system if the data upon which the WPB depended were to be used to the greatest advantage. Time and again the WPB had to adjust basic control procedures to fit the realities of sound data-collection methods. This does not mean that statistical methodology dictated either organization or control policy. Rather it means, as the WPB discovered, that since all three had to be related as closely as possible to the fundamental methods and procedures of industrial operations, all three had to be coordinated for the most effective operation of each.

Finally, as the WPB also discovered, the best statistical system will fail to reach its potential if there is not a free flow of information among all levels of management and if management at any important level fails to admit the importance of accurate factual knowledge in making decisions and administering them. An emergency statistical reporting system, even though it has the characteristics noted here as being important, will be no better than the degree of coordination of other elements of management and the intelligence with which it is operated and with which its data are used.

Applying the Technical Methodology of a Reporting System

In striving to produce the statistical discipline which it came to recognize as the basis of its operations, the WPB met with every important data-collection problem imaginable. In the solution of these problems it experimented with almost every known technique. Selected technical lessons which the war period taught are the substance of the preceding articles. Since the discussion is a concentrated condensation of a vast experience, it is not desirable to attempt further summarization at this point. A better purpose can be served by attempting to set forth a few of the more important principles which experience has taught should be kept in mind when applying the highly technical methodology of industrial factual accumulation in emergencies.

Experience shows that it is both necessary and possible to create a coordinated emergency industrial reporting discipline. An emergency industrial control agency is in a real sense an operating holding company over the productive machine of the nation. Such an agency

must be in a position to collect accurately and quickly at one central point or points summaries and details of the millions of actions which are taken in the highly complex production machine. The WPB's experience shows that it is possible to draw the factual threads of industrial action together at a central point or points in such a manner that the data are well coordinated, comprehensive in character yet penetrating in detail, accurate, and quickly tabulable for prompt use. If methodology is well planned, carefully applied and used with intelligence it can be made flexible enough to provide data to meet all important problems. The WPB never created such a statistical discipline but its experience leaves no doubt about both the need for it and the ability of government and industry to create it.

The development of a coordinated, flexible and useful reporting system is impossible in an emergency without some centralized control of the issuance of questionnaires and the application of technical methodology. In time of crises every pressure on administrators is toward immediate action. The drive for facts is so great that any promising method is attempted. The results of unbridled issuance of questionnaires in an emergency cannot be less than an unjustified burden on industry and chaotic government administration. By centralizing control over questionnaire issuance a focal point is provided for gathering together all sorts of questions relating to questionnaire policy and methodology. An opportunity is provided for such questions to be resolved in the light of the best policy and methodology. If this function is managed with wisdom the questionnaire system and its techniques may develop into and operate as a balanced fact-gathering mechanism. If it is not, the best plans and intentions will not create a coordinated data-collection system.

To approach a coordinated statistical methodology the many technical principles composing it and inherent in its satisfactory maintenance and operation must be well-formulated and clearly set forth. A balanced reporting system represents a highly complex structure in which details are of great importance. Not only may injudicious instructions, or poorly-phrased questions leave management without facts for crucial decision, but inattention to minute household details relating to such matters as mailing lists, follow-up procedure, editing, coding, and machine tabulation may also invalidate an important survey. And, paradoxically, the more complicated in substance and comprehensive in coverage the questionnaire the more important become these details to valid results. Questionnaires of small magnitude may be hand-processed to avoid administrative problems in govern-

ment. Comprehensive questionnaires cannot be so pampered. Slight error or miscalculation in any technical detail involved in their life cycle may become so magnified that the final results may be either suspected or ignored by management. The existence of a set of carefully considered technical standards which may be uniformly applied if the will and machinery exist to do it is a primary requisite to statistical discipline.

One of the most important groups of statistical standards which must be created and applied with the greatest skill is that relating to technical nomenclature and statistical units of measure. Not only are definitions and classifications of products, materials, plants, etc., crucial in the good use of data but in their determination there are present many problems which are administratively complex. Units of measure are, of course, basic in the functioning of a reporting system. Their determination and application also present exceedingly difficult and complicated problems. Statistics are a method of communication. They are, in addition, instruments of policy and administration. They are useless to serve these basic functions unless they are founded on a clear, understandable and administratively usable language of definition, classification and units of measure.

The WPB found that it is not important nor always desirable to embrace the complete universe of a subject matter in order to produce meaningful data. Concentration in industry of metal consumption, component usage, manpower requirements, and other production elements, is such that facts necessary to exercise control can be acquired from much less than the totality of units in a given area. For almost every product and material not more than 15 per cent of the total possible plant respondents account for not less than 85 per cent of the total economic activity for that product or material. By limiting reporting to these large units, and by exercising control in smaller units by means of general regulations, the objectives of emergency industrial control can be attained without imposing serious record-keeping burdens on the bulk of the small enterprises and serious clerical problems on government.

Reporting methodology must be geared to industrial practices if prompt and accurate data are to be acquired. The relationship between the design of centralized control procedures and methods of industrial operation must be carefully considered when developing administrative techniques. Experience teaches that any control procedure which accommodates itself to industrial practices can be instituted with a minimum of operational friction. Every new control system must go

through a warm-up period to familiarize industry with the new routines and to permit necessary adjustments to them. To the extent that operational changes are held to a minimum, this initiation period is shortened. In addition, any control system which is based upon industry's existing record-keeping practices is bound to yield better and more complete results. The closer the reporting framework and the controls which it supports approximate industrial practices, the quicker and easier it is to attain the objectives established for controls.

Better results are obtained if reporting is on a *quid pro quo* basis. Despite the obvious load which some complicated questionnaires, such as allocation and scheduling forms, imposed on industry, particularly on those plants whose record-keeping systems were not designed to yield readily the required data, reports were received quicker and with more accuracy than most experienced statisticians thought possible. On the other hand, many comparatively simple questionnaires designed to acquire general information for broad policy analysis were often met in industry with intense and bitter opposition. Response was slow and replies incomplete. The basic reason for this paradox is simple: the former eventually resulted in a right for the respondent to do something; the latter produced no direct advantages to the respondent. The obvious lesson to be drawn from this experience is that the more closely statistical data requirements are tied to application-type reporting forms the better the data will be both with respect to accuracy and completeness and the quicker returns will be made by respondents. In addition, industrial criticism and objections to a questionnaire, other things being equal, tend to decrease with the importance of the privileges or rights bestowed through the reporting instrument.

A comparatively long period of time is required to produce accurate and complete data through a new form. On the basis of the WPB's experience, from three to six months are required from the first submission of a new complicated reporting instrument to the time when both industry and government function reasonably smoothly on the basis of it. With advance publicity this period can be reduced, provided the questionnaire is not too complicated nor has in it injudicious procedures or questions difficult or impossible for industry to answer.

It is possible to reduce the number of data requests if a case analysis is made of the internal use of the data derived from the questionnaire and adequate provision is made within the control organization for the free and prompt exchange of information. The basic categories of information required for emergency management are not large. If sound technical standards relating to such matters as units of measure, no-

menclature, time periods and other details are carefully applied, and if a well-planned statistical system is used to support intelligent and suitable control policy, there is no reason why a relatively few reports cannot be made to satisfy fully management's requirements for facts. The experience of the WPB adequately supports this conclusion.

The adaptation of accounting techniques to statistical methods can be made effectively both to support implementation of policy decisions and to provide the necessary factual data for review of past decisions and new determinations. A statistical-accounting technique can exist side by side with a statistical reporting system and can constitute an important part of the entire data-collection framework. Although the WPB's material accounting was not as accurate as financial accounting, it performed for production control much the same functions as financial accounting performs for industrial financial control. By introducing tolerances permitted by material accounting needs, accounting control was instituted and operated at a far less expenditure of time and energy than financial accounting. Statistical accounting is a practical and necessary part of any emergency data-collection structure used to support control policy and its administration.

An adequate reporting system cannot be devised overnight. In early 1940 a tremendous mass of information was available in Washington pertaining to the operation of the industrial system. The data, however, could not be coordinated nor organized to present a comprehensive statement of resources. Data on potential war demands were for all practical purposes non-existent. The concept of an integrated emergency reporting system was unborn. Developing the concept and nursing it to maturity was a difficult and time-consuming task.

Experience of the WPB shows that it never fully recovered from these early statistical deficiencies. Makeshift schemes and organizations became deep rooted. It was impossible to erase them completely. As the war progressed the value of satisfactory new control systems were jeopardized because of existing inept procedures originated at a time when a scramble for factual knowledge resulted in the introduction of any procedure which appeared capable of producing some results. The erection of an effective statistical reporting system was, therefore, doubly handicapped.

Our experience after 1938 now demonstrates that we were as ill-prepared to create and use the proper techniques for collecting information as we were to wage modern war and produce quickly the weapons needed for its prosecution. Although we had a larger peacetime statistical organization than any other country in the world, we were not

prepared to provide the facts for an all-out economic effort. The reporting system merely provided a foundation. It was comparable in this respect to the world's largest steel capacity and the world's largest industrial capacity. We had the resources; we did not have the end-products that we needed.

War, we found, required not merely general facts; it demanded that facts be obtained on a broader basis, on a wider variety of subject matter, and in a shorter period of time than had ever before been necessary. To do this job we built upon prewar experience and techniques. Frequently, the needs of the emergency required substantial changes in available methodology. In some cases, emergency developments moved so far beyond original techniques as to almost preclude recognition of the foundation upon which they had been built.

For years we had collected biennially a *Census of Manufactures* which covered every fabricating establishment in this country. Data were collected for each establishment showing its employment, the products which it made, its consumption of electrical energy and the dollar value which it added by manufacture. The census was taken in odd-numbered years and became available to the general public some two years after the end of the period which it covered. In addition to this periodic census, we also had a large variety of current reporting series in the industrial field. Some of these were developed by the Federal government. To a larger extent they were the product of trade associations and business groups which cooperated in the collection of certain data necessary to their current operations. Over the years the Census data had provided a series of benchmarks against which current reports could be measured and projected. These were valuable statistics. For the purpose for which they were designed and used, they were undoubtedly all that was necessary.

The methodology and techniques were also adequate for the purposes for which peacetime industrial data were collected. In peacetime, data were collected to permit government and business administrators to make broad general judgments. Seldom were they used to administer specific actions. Business and government operated on the basis of a multitude of individual choices not directly related to and administered in conformance with national policy objectives. In wartime, the direction of the economy was founded on a single policy objective made by government. And, what is perhaps more important, the implementation of that specific central judgment depended on directions issued by the central government. Under such circumstances statistics became the basis of action. The data therefore had to be accurate and up-to-the-

minute. As a result, we could no longer operate with the methods of peacetime. A methodology had to be found by means of which penetrating industrial data could be collected in the most accurate and comprehensive terms and made available to administrators very quickly.

Our wartime experience demonstrated that no matter how urgent the need or how willing the cooperation, an integrated reporting system suitable for emergency management cannot be improvised or built in a short period of time. The development of useful information is a product of knowledge, skill, experience, and time.

If we are to profit from this experience and avoid costly errors in a future emergency, we must now determine upon a course of action which will permit us to preserve the accumulated knowledge of the past and to use it in molding a better system for future periods of crisis. We must develop not only the blueprint of the reporting system needed in an emergency, and relate it closely to emergency control plans, but we must insure that the major features of the statistical system are incorporated as far as possible in the current statistical reporting structure of government. A relatively small expenditure now to provide for emergency statistical preparedness, and if need be continued for the next 40 or 50 years, would be minute compared with the costs incurred in the three years of confusion attending the recent war effort, or the early period of confusion which we can be sure will be part of any future emergency program in the absence of a continuing pre-emergency effort to prepare an integrated emergency industrial reporting system.

BOOK REVIEWS

Edited by
OSCAR KRISEN BUROS
Rutgers University

Practical Business Statistics, Second Edition. *Frederick E. Croxton* (Professor of Statistics, Columbia University, New York, N. Y.) and *Dudley J. Cowden* (Professor of Economic Statistics, University of North Carolina, Chapel Hill, N. C.). New York 11: Prentice-Hall, Inc. (70 Fifth Ave.), 1948. Pp. xix, 550. Text edition, $6.35; trade edition, $4.75.

REVIEW BY ALFRED CAHEN
Textile Economist, Business Information Division, Dun and Bradstreet Inc., 326 Broadway, New York 8, N. Y.

THIS up-to-date revision of the popular original 1934 edition offers in understandable form some of the newer techniques on sampling, quality control, and time series analysis. Particularly effective are the step-by-step clear-cut solutions of numerous problems selected from business and government statistics. It remains one of the best texts on business statistics.

Some specific contributions notably well presented include Chapter 3, "Statistical Tables," and Chapter 4, "Rates, Ratios, and Percentages." This latter chapter effectively points out the inter-dependence of statistics and accounting. Chapters 5, 6, and 7 deal thoroughly with graphic analysis.

Time series analysis is effectively presented in Chapters 11, 12, and 13. Some simplified techniques are particularly useful as preliminary adjustment of data for price changes, calendar variation, and population changes.

Chapter 14 on correlation is perhaps oversimplified. The illustration on relationship between hardness and tensile strength of 27 pieces of wrought aluminum alloy results in a coefficient of +.991. The authors indicate in the preface that Chapters 1 to 15 may be used for a one-semester course and Chapters 16 to 22 for the second semester. Multiple correlation is dealt with six chapters later in Chapter 20. A realistic question arises whether presentation of simple correlation alone (separate from multiple) may not overimpress the student. "A little knowledge is a dangerous thing."

Chapter 16 deals with the normal curve. Happily the authors refer only once to coin tossing and not at all to dice or cards. These three illustrations are generally a curse upon both teacher and students of business statistics as they depart so far from reality. One example used to illustrate the normal curve is life experience of wooden telephone poles. Few business series can be fitted by a normal curve. It was originated by astronomers and generally applied to physical and biological data. The family of curves of the Pear-

sonian or Gram-Charlier series cannot be studied in a first year course. Therefore a chapter on the normal curve alone probably contributes little that a business statistics student will apply later in practice. However, the normal curve is important in the study of sampling distributions.

Some current indexes are analyzed in Chapter 21, consisting of the principal governmental price and production indexes. This chapter is simply and interestingly written. Perhaps it could best follow immediately Chapter 15, "The Construction of Index Numbers." The allocation of only one-half page to the logistic and Gompertz curves hardly seems adequate.

The final chapter, "Budgeting and Forecasting," is effectively presented except that no mention is made of unfilled orders and inventories—both of which are rather frequently used in forecasting.

The book contains 14 appendices: including mathematical tables as values of t, F, χ^2, etc; useful reference tables on sums of the first four powers of the first 50 natural numbers; squares, square roots, and reciprocals; logarithms; and interesting side lights such as rounding numbers and flexible calendar of working days.

So much for the high spots of the book. With so many worthy contributions, one may ask what are the limitations? That concerns primarily the use of the word "practical" in *Practical Business Statistics.*

Practical is defined by *Webster's International Dictionary* as "of, pertaining to, or consisting or manifested in, practice or action."

The emphasis in this text is primarily on techniques as indicated by the large amount of space devoted to the following topics:

SUBJECT	CHAPTERS	PAGES
Time series analysis	3	86
Correlation	2	54
Reliability and tests of significance	2	49

In contrast the authors are parsimonious on problems which their students will most likely encounter later in jobs in practical business statistics. Some of the larger requirements of the operating statistician are in the following fields (Note the number of pages given by the authors.):

SUBJECT	PAGES
Sources of data	8
Detection of errors	5
Questionnaires	5
Sample design (not including chapters on tests of significance)	7
Collection of data by interview or mail	2
Editing	1
Tabulation methods	6
Graphic presentation (adequately covered)	52
Report writing	—

These are realistic problems which both young and old practicing statisticians grapple with day by day.

Sources of data are the first step in tackling a business problem. The

authors do provide in Appendix A a five page list of general sources on production, price, wage and similar data. A concrete problem showing a student how to find the pertinent information on one industry such as wood furniture might be more constructive.

Detection of errors is probably the most important single problem which the practicing statistician encounters every day of his working life. The various checks and counter checks to help detect errors are scarcely mentioned in this text.

The selection of the proper tabulation method—hand, machine, or semi-mechanical—is highly important to the practicing statistician both in time and cost. Proper tabulation method sometimes saves several thousands of dollars on a single project. The six pages devoted to this subject by the authors oversimplify the problem by stating mechanical tabulation is desirable when a large number of schedules is being tabulated with many entries for numerous tables. This resembles the sales talks of the representatives of the International Business Machines Corporation and the Remington Rand Company. The determining factor in the use of mechanical tabulation is not the size of the project. When information needs to be both multiple sorted and repeatedly readded, then mechanical tabulation offers operating economics.

However, if it is largely a problem of multiple sorting without summations, the Keysort or EZ Sort systems can readily contribute less costly operation than mechanical tabulation for the same results. The reviewer saw an operation of Keysort at the Mutual Life Insurance covering 1,250,000 accounts—certainly a large scale operation. The authors do not mention the EZ Sort system which has a patented advantage of five holes to the inch compared with four holes for the older Keysort system.

In tabulations, regardless of size, where summation is the primary problem with little or no sorting, then the peg board system and the accounting machines of National Cash Register Company, both offer economies compared with mechanical tabulation.

Appendix B describes several calculating and adding machines. Crelle's tables offer a very economical means of multiplication, but are not mentioned. Although three different calculating machines and one adding machine are pictured, no reference is made to the comptometer. Various tests have shown that adding a set of figures twice on the comptometer is both more speedy and accurate than adding once on the adding machine and then reading back the tape. Tape reading involves the human frailty of an individual's mind wandering instead of listening to figures. Hence errors are not detected.

The authors have contributed much technical perfection and pedagogical skill in the revision of this volume. Some additional time could have been profitably spent to find out how their former students are actually putting statistics to use in practical business.

Dr. Juan Kimmelman rather aptly phrased the problem in the June 1944

Journal of the American Statistical Association "In many cases, after having devoted several years to practical work, even in high positions, he will admit that 80 per cent of the knowledge which he obtained in the university represents dead weight for him, and that it is neither needed nor of any possible use.

The examples in this text are selected from actual operations. However, the sources of these illustrations are largely from governments, institutions, and quasi-public giant corporations, the stratosphere of American business. They are *technical rather than typical* problems encountered from day to day by the rank and file of business statisticians.

All of the foregoing comments can probably be boiled down to a matter of relative emphasis. Formula problems are an essential part of a first-year statistics course. However, problems of collection of data, detection of errors, tabulation, and report writing are never given sufficient space and emphasis in a first-year statistics course.

The reviewer recommends this volume as a handy guide that every statistician will like to have on his desk for reference. Altogether it is an excellent text on business statistics—practical or otherwise.

Mathematical Methods for Population Genetics. *Gunnar Dahlberg* (Head of the State Institute of Human Genetics, Uppsala, Sweden). New York 3: Interscience Publishers, Inc. (215 Fourth Ave.), 1948. Pp. vii, 182. $4.50.

REVIEW BY HOWARD LEVENE
Instructor in Mathematical Statistics and Biometrics
Columbia University, New York, N. Y.

OUR present knowledge of genetic phenomena in human populations is still in a comparatively rudimentary state, and any advances are likely to prove of great benefit to the human race. The problems involved are: (*a*) to learn by investigation what sort of assumptions may reasonably be made about the processes which occur; (*b*) to formulate precise mathematical models embodying these assumptions and deduce their consequences; (*c*) to estimate the parameters involved. The Scandinavian countries, with relatively homogeneous populations, excellent vital statistics, compulsory reporting of hereditary defects, and governmental interest in the problem, offer a particularly good opportunity for investigation of problems (*a*) and (*c*). The author, who is Head of the State Institute of Human Genetics in Sweden, has played an important part in investigating all three problems. He has written a work with many excellences, combined with some defects.

This book is a translation, with minor changes, of the author's "Mathematische Erblichkeitsanalyse von Populationen," *Acta Medica Scandinavica*, Suppl. 148, 1943. The author has deliberately omitted all discussion of the work by the Fisher school and others on estimation of mode of inheritance, linkage relationships and gene frequencies for particular characters; and has made little use of the investigations of Wright, Fisher, Haldane and others

on evolution in natural populations. Within the author's chosen field, the book suffers from the defects to be expected of a book written on an essentially mathematical subject by a non-mathematician. Most of the proofs involve verbal reasoning and are consequently hard to follow and check. Furthermore, the precise mathematical model involved is not always clearly evident. In particular, the reviewer has felt for some time that geneticists use the term random breeding to cover a number of different processes that may lead to different results in finite populations, and the author's discussion seems to have this same fault. The author also fails to distinguish between the actual values of gene frequencies and their expected values, although at other times he makes effective use of the concept of random fluctuations in small "isolates." Furthermore, the modern theory of statistical inference is virtually ignored. In the section "Isolates and Racial Differences" a type of discriminant analysis is introduced which is in general inefficient in comparison with that of Fisher.

The above criticism should not be taken to imply that this is not a valuable book. For biologists and men of affairs it discusses clearly the problems of heredity in human populations, and shows the importance of further research along these lines. The final chapter in particular should serve as a valuable antidote to the more rabid eugenists and racists. (The author states that the first edition was intended in part for the post war edification of the German people.) The wealth of tables and graphs interpreting the various formulas should be particularly welcome to the less mathematical reader, and the conclusions drawn from them are likely to be at least qualitatively correct.

For the mathematical statistician and the biometrician, this book will be primarily valuable as an introduction to a new discipline of great practical importance which furnishes some problems of no mean mathematical difficulty. The reinvestigation and extension of the author's results, using the theory of stochastic processes and modern methods of statistical inference, provides a wealth of problems at many different levels of difficulty. All things considered, it is a book which will well reward reading by those with any curiosity toward this subject.

Principles of Biological Assay. *C. W. Emmens* (Head of the Department of Veterinary Physiology, University of Sydney, Sydney, Australia). Foreword by Percival Hartley. London W. C. 2: Chapman & Hall Ltd. (37 Essex Street), 1948. Pp. xv, 206. 21*s*.

REVIEW BY LILA F. KNUDSEN
Statistician, Food and Drug Administration
Washington 25, D. C.

UNTIL just recently, there has been no one comprehensive book on statistics in biological assay though the need for it has been great. There have been innumerable articles in scattered journals since most of the developments in this field have been in the last fifteen years, but nowhere have the

various statistical technics been gathered in one book. *Principles of Biological Assay* compares with an excellent recent book entitled *Probit Analysis* by Professor D. J. Finney of Oxford, which covers the field of biological assay with special emphasis on quantal (or percentage) responses and gives a rather thorough background in design and interpretation of such biological assays. Dr. Emmens' book attempts to cover a wider approach and includes quantitative as well as quantal responses on biological assays with one chapter devoted to the work of Finney and Wood on microbiological assays.

It seems that the title *Principles of Biological Assay* is misleading and should have been *Statistics in Biological Assay*, or maybe *Statistical Design and Analyses of Biological Assay*, since the entire book deals with the statistical design and analysis and not with the pharmacological aspects of assays as do other books in this field, such as J. H. Burn's *Biological Standardization* or Katherine Coward's *The Biological Standardization of the Vitamins.*

Dr. Emmens is primarily a pharmacologist who has found statistics extremely necessary in arriving at sound conclusions in experimental biology and the evaluation of strength and comparative effects of drugs. In 1939 he wrote a report on biological standards entitled "Variables Affecting the Estimation of Androgenic and Oestrogenic Activity" which was issued by the Medical Research Council of Great Britain. Dr. Emmens has come a long way, statistically speaking, since writing that report and his present book covers some of the newer important developments in statistical evaluation of biological assay such as Irwin's contribution for taking into consideration the error of the slope in calculating the fiducial limits of error, and Finney's probit plane analysis.

The book touches on the usual introductory statistical concepts; deals rather completely with dosage-response lines; describes various experimental designs (Latin squares, balanced incomplete blocks, etc.) that have been used in biological assay; gives formulas for calculating potency, standard errors, fiducial limits, combining results of several assays, chi squares, etc. However, emphasis seems to be on the manipulation of numbers rather than the assumptions made and the interpretations to be given the results. Even at that, the worker in the biological laboratory is given no inkling that in many cases it is possible to greatly simplify the formulas for routine calculations. The organization of the book seems a bit jumbled to the reviewer. The twenty chapters are each very short and each section does not seem to lead logically into the next section. The two chapters devoted to design of experiments are widely separated. Various tests of significance are scattered through the book. An entire chapter is devoted to a rather hazy explanation of polynomial coefficients and when factorial coefficients are introduced three chapters later, the author does not mention that they include the impressive polynomial coefficients. Student's t test occupies one paragraph and the only formula given is that for comparing a sample mean with the population value of the mean. Though the symbol t is used later in comparing two slopes and in calculating fiducial limits of the potency, nowhere in the book is an ex-

planation given of the assumptions underlying it and the way it should be interpreted.

The section on degrees of freedom seems unnecessarily complicated for absorption by a pharmacologist. The author apparently is trying to lead up to analysis of variance and factorial coefficients in two pages but simply confuses the reader.

Terminology can be very confusing. Dr. Emmens probably didn't realize that he gives three widely different concepts to C on pages 84, 86, and 105. It would have been possible to choose some other letters or symbols for two of the concepts. Another thing that may be confusing is the omission of the usual parentheses after the summation sign S in many equations.

It is rather surprising to see a book of this kind without lists of references for additional reading. It must have been an oversight that no mention is made of many of the numerous articles and statistical texts that would add to the reader's comprehension of the subject; although full credit is given in the text for examples cited and most formulas quoted. The method for evaluating biological assays given by E. B. Wilson and Jane Worcester in a series of papers published in the *Proceedings of the National Academy of Sciences* in the spring and summer of 1943 is not mentioned nor is Epstein and Churchman's "On the Statistics of Sensitivity Data" in the March 1944 issue of *Annals of Mathematical Statistics* nor Berkson's "logits."

It is consoling to find the statement on page 188 "The weighted mean log potency should have a variance $V\bar{M}$ given by

$$\text{(c)} \qquad \frac{1}{V\bar{M}} = S\frac{1}{VM}$$

which should not differ significantly from the alternative estimate

$$\text{(d)} \qquad VM = \frac{Sw(M - \bar{M})^2}{Sw(r-1)}, \quad \text{where} \quad w = \frac{1}{VM}."$$

and that "the best estimate of the potency available remains the weighted mean, which has a variance given by equation (d)." The implication is that even if the state of statistical quality control is known to exist the internal variance of an assay may give too small an error of the combined potency estimate. The left hand side of equation (d) contains an obvious typographical error. In the book it is written as VM instead of $V\bar{M}$. The denominator of the right hand side of this same equation would be clearer if written $(r-1)Sw$.

On the whole, this book will be a welcome addition to the library of a pharmacological laboratory, particularly as a source book for formulas, and very helpful in promoting the use of statistics in biological assay.

Traffic Performance at Urban Street Intersections. *Bruce D. Greenshields, Donald Schapiro,* and *Elroy L. Ericksen.* Yale Bureau of Highway Traffic, Technical Report No. 1. New Haven, Conn.: Bureau of Highway Traffic, Yale University, 1947. Pp. xv, 152. Gratis.

REVIEW BY HENRY K. EVANS

Highway Specialist, Chamber of Commerce of the United States, Washington, D.C.

THIS 148-page technical treatise is a highly theoretical discussion of vehicular traffic movement characteristics (primarily at street intersections), comparing results of empirical field observations with mathematical calculations. The field studies employed a specialized photographic technique; the theoretical calculations revolve about the use of the laws of probability, especially the Poisson theory.

It is found that the field observations of vehicular spacings and groupings, (in moving traffic) agree very well with mathematical estimates of these characteristics. Therefore support is gained for the use of the mathematical approach in estimating traffic behavior, and the authors devote half of the book primarily to applying the laws of probability to different hypothetical traffic problems. The proof of the successful application of the mathematical method of calculating behavior patterns is well substantiated by the results shown in the book. However, as a useful tool to the practising traffic engineer or planner, the methods shown have no immediate practical value, principally because such precise knowledge of these behavior patterns, no matter how gained, bears negligible relationship to traffic engineering techniques commonly employed.

For example, various mathematical methods of calculating vehicle seconds delay at street intersections or drawbridges are presented. The traffic engineer is not as interested in knowing the precise amount of delay as he is concerned with knowing whether or not improvements in traffic control result in substantial reduction in over-all delay, which is easily observed by before and after field observations and simple arithmetic calculations. (The field observations would be necessary whether or not the mathematical refinements were applied.) Not enough is known about economic costs of delay to warrant any high degree of exactitude in determining vehicle seconds delay.

The highly theoretical aspect of the book is illustrated by the solution of a given problem of determining whether or not the frequency of obscuring vision (one driver's vision of a traffic island obscured by any vehicle within a 75-foot distance ahead) would be doubled by doubling the traffic volume. Use of Poisson's theory develops the estimate that 16 per cent of the drivers would experience obscured vision where traffic volume averaged 300 vehicles per hour, and that the per cent would be 29 per cent at double the volume. As a mental calisthenic this method bears some merit but has no practical use.

Another similar application of the laws of probability is made to a hypothetical traffic signal timing problem. Here the object is to find the shortest signal cycle length at an intersection which will pass the entering traffic volumes with failure to clear out the waiting vehicles only 5 per cent of the time. The method is useful primarily and only for the insight it gives into the traffic movement capacity of a roadway as limited by signal cycle length. Practically there are so many other factors bearing on selection of optimum signal cycle length that the theoretical method has little practical use.

Application of the Poisson theory to estimate the bunching of vehicles arriving at a parking lot or garage is brought out. This has a value to designers in making it possible to estimate the reservoir space required to care for vehicles arriving faster than they can be parked and is one of the most valuable applications shown of the mathematical approach to practical problems.

Whereas the techniques illustrated in this book have little application in solution of practical traffic problems, they do constitute valuable tools for research into generalized warrants for application of traffic control devices such as stop signs and traffic control signals. As yet the traffic engineering profession has not reached a verdict on what volumes of traffic or frequency of accidents justifies installation of stop signs at an intersection. The method shown of computing the percentage of unnecessary stops at a stop sign suggests that this knowledge might be used to determine that below certain levels of traffic volumes on two streets the percentage of unnecessary stops would be high enough to justify declaring that stop signs were unnecessary and a nuisance (provided other factors such as traffic accidents and view obstructions did not require such signs).

Although this book provides no ready answers or useful tools for the traffic engineer, it deserves to be acclaimed a milestone in the progress of traffic engineering research. It is the most scholarly and scientific investigation into the characteristics of traffic flow patterns yet published.

The primary contribution to the traffic engineering profession is the proved application of Poisson's probability formulae to analyses of traffic flow. Whereas it is doubtful that such application will prove of every day use for individual problems in setting signal cycle lengths, determining need for stop signs and other traffic facilities, or comparing relative efficiencies of different control devices, because of the many other non-mathematical factors affecting traffic flow controls and facilities; it is felt that the mathematical tools set forth in this report will be of inestimable assistance in performing basic research for the purpose of establishing general standards and warrants for different methods of control, sizes of traffic facilities, etc.

This method of traffic engineering mathematical analysis is akin to the first mathematical studies of nuclear physics. Whereas it has negligible immediate practical application, it is the entryway into a better understanding of the principles involved in traffic movement, which will inevitably result in practical application.

Theory of Probability, Second Edition. *Harold Jeffreys* (Plumian Professor of Astronomy, University of Cambridge, Cambridge, England). London E.C.4: Oxford University Press (Amen House, Warwick Square), 1948. Pp. viii, 411. 30*s*. (New York 11: Oxford University Press [114 Fifth Ave.]. $9.00.)

REVIEW BY HERBERT ROBBINS

Associate Professor of Mathematical Statistics
Institute of Statistics, University of North Carolina
Chapel Hill, N. C.

THE SECOND edition of this work on statistical inference differs from the first mainly in the addition of a theory of invariance aimed at establishing the logical consistency of the author's rules for setting up prior probabilities. The whole theory, based on Bayes's principle of inverse probability, lies entirely outside the modern development of statistical inference in the work of Neyman, E. S. Pearson, and Wald. Its relation to Fisher's work is more obscure; according to Jeffreys his results are in general agreement with Fisher's, although the underlying reasoning is not always the same.

Perhaps the most striking feature of Jeffreys' work is that his methods for testing hypotheses or estimating parameters take no account of the cost to the experimenter of accepting a false hypothesis or using an erroneous estimate. For example, a null hypothesis is to be rejected when the ratio of its posterior probability to that of the composite alternative is too small, the prior probabilities being determined by a set of conventions into which the cost of making errors does not enter. This procedure, which is based on an axiomatic concept of probability that has nothing to do with relative frequency, is not supported by any operational justification. Within Jeffreys' theory one cannot ask, let alone answer, whether the procedures are better or worse as operating rules than others which might be put forward as alternatives. This was indeed the situation in statistical inference thirty odd years ago.

Whatever one's views on the foundations of statistical inference, one cannot fail to profit by the mathematical ingenuity, keen physical intuition, and common sense which Jeffreys brings to bear on a wide variety of practical problems. Statistical theory will continue in large measure to find the inspiration and motivation for its real advances in the practice of the best statisticians. The appearance of the second edition of a book as provocative and rich in content as Jeffreys' *Theory of Probability* is a welcome event.

Rank Correlation. *Maurice G. Kendall* (Joint Assistant General Manager and Statistician, Chamber of Shipping of the United Kingdom, London E.C.3, England). London W.C.2: Charles Griffin & Co. Ltd. (42 Drury Lane), 1948. Pp. vii, 160. 18*s*.

REVIEW BY E. J. G. PITMAN
Professor of Mathematics, University of Tasmania
Hobart, Tasmania

IN THE words of the preface: "Until a few years ago rank correlation was a rather neglected branch of the theory of statistical relationship. In the practical field it was generally regarded, except perhaps by psychologists, as a makeshift for the correlation of measurable variables; and in the theoretical field it seemed to present no interesting or important problems. That situation has changed. Practical applications of ranking methods are not only being extended in psychology and education but are being made in other subjects such as industrial experimentation and economics. The theoretical properties of order-statistics have received much attention and are throwing important light on some difficult questions of statistical inference."

As the title indicates, this book deals only with the application of ranking methods to testing independence and to estimating degree of dependence of chance variables. The basic problem is, given n pairs of values of the chance variables X and Y, to decide whether X and Y are independent or not. By replacing the observed X values by their ranks, and the Y values by their ranks, we are able to test the hypothesis of the independence of X and Y without making any assumptions about the forms of the distributions. Moreover, we can do this when the values of one or both variables cannot be properly measured but can only be arranged in order of magnitude.

The book gives a clear and comprehensive account of the properties of Spearman's coefficient of rank correlation ρ, and of the coefficient τ, which depends essentially on the number of inversions in the ranks of one variable when the ranks of the other variable are arranged in natural order. Neither here nor elsewhere has the latter coefficient been given a name in English. The complications arising from ties are thoroughly treated.

The problem of m rankings, and partial rank correlation are dealt with, and also the relation of ρ and τ to the population correlation coefficient in samples from a normal bivariate population. The last two chapters discuss paired comparisons, as when an observer compares n objects two at a time without necessarily in the end arranging the n objects in a linear order. The tables necessary for making all the tests described in the book are provided.

Lack of precision in statement and in notation, and insufficient explanation sometimes make the author's argument difficult to follow. For example, on page 58 we have

$$E(c^2) = E\Big\{\sum (a_{ij}b_{ij})\Big\}^2$$
$$= E\Big\{\sum (a_{ij}^2 b_{ij}^2) + \sum{}' (a_{ij}b_{ij}a_{ik}b_{ik}) + \sum{}'' (a_{ij}b_{ij}a_{kl}b_{kl})\Big\}$$

where $\sum'$ denotes summation over values for which $j \neq k$ and $\sum''$ over values for which $i \neq k$ and $j \neq l$.

The summation in the first line is intended to include only terms in which the two first suffixes are the same or tied, and the two second suffixes are tied, i.e., terms like $a_{ij}b_{ji}$ are not to be included. It is natural to assume that the same convention applies to the first summation in the second line, and this is in accordance with the statement about $\sum''$. The author then goes on to show that the mean value of $\sum''$ is 0; but this requires $i \neq k$, $i \neq l$, and $j \neq k$, $j \neq l$. If that is so, the first summation must include terms in which the b^2 suffixes are the same as the a^2 suffixes, but in reverse order, so that terms like $a_{ij}{}^2 b_{ji}{}^2$ are to be included. (Note: $a_{ji} = -a_{ij}$, $b_{ji} = -b_{ij}$.)

Again, on page 61 it is stated that when summations extend over values of suffixes which exclude certain values (e.g., $i \neq k$) we can replace them by complete summations. This will puzzle the conscientious reader, as no reason is given. What is meant is that we can do this without affecting the dominant terms of their mean values. This is explained later, page 66, in the discussion of a more general case.

It is an excellent idea to publish a monograph on a selected portion of statistical theory, and it is to be hoped that other workers in statistics will follow this author's example.

Report on the Scheme for the Improvement of Agricultural Statistics. *V. G Panse.* Imperial Council of Agricultural Research. Delhi, India: Manager of Publications, 1946. Paper. 3*s.* 9*d.*

REVIEW BY S. LEE CRUMP
Iowa State College, Ames, Iowa

THE BROCHURE under review is arranged in eight chapters and four appendices. The main chapters discuss the following topics: statistics of area under agricultural crops, crop cutting experiments, forecasts and estimates of yield, sampling surveys for estimating yields. The remaining chapters are of a more general or incidental nature.

The total production of a crop in India is currently estimated from the following formula: Area × Normal Yield × Seasonal Factor. Dr. Panse has approached the problem of improving agricultural statistics with this formula as a take-off point. Each of the first two factors is discussed in considerable detail while the third is treated briefly.

The area under agricultural crops is assumed to be known with a high accuracy since an annual census is taken. The chief problems in this connection are those arising from mixed crops and from divided fields. Suggestions are made for reducing the error from these two sources. It seems to this reviewer that Dr. Panse may underestimate the importance of other errors which seem always to introduce incompleteness in censuses.

By far the greater part of the discussion is devoted to the problem of estimating yield. Dr. Panse has undertaken a thorough study of the methods of

estimating yield in current use. They all suffer from subjectiveness at almost every stage. The crop cutting experiments which should provide the best estimates of yield are not properly conducted, and are frequently ignored. Throughout the discussion the point is made that "normal yield" is an ill-defined concept and should be abandoned in favor of estimating the yield directly each year.

Finally, the results of a random sampling survey for estimating the yield of wheat are presented and discussed. As has always been the case, the random sampling survey method seems to offer the only economical scheme for obtaining reliable estimates.

The first three appendices give the details of the operation of the sample survey. In Appendix 4 two papers by V. G. Panse and R. J. Kalamkar on sample surveys are reprinted.

Dr. Panse should be commended for his approach to the problem. At each stage of the investigation he has attempted to suggest improvements in the estimates within the existing framework. That the final results indicate the overwhelming superiority of random sample surveys is not because Dr. Panse has arbitrarily rejected the existing scheme for making estimates.

Methods of Estimating Vital Statistics of Fish Populations. *William E. Ricker* (Professor of Zoology, Indiana University, Bloomington, Ind.). Indiana University Publications, Science Series No. 15. Bloomington, Indiana: Indiana University Bookstore, 1948. Pp. v, 101. Paper. $2.00.

REVIEW BY CHARLES M. MOTTLEY

Operations Analyst, Headquarters United States Air Force
Washington 25, D. C.

THE VITAL statistics of fish populations have undoubtedly been intriguing from the earliest times. The first fishermen must have wondered why the fish were abundant at one time and not the next. The surface of the water presents a barrier which man can seldom penetrate. The human census taker can draw representative samples, interview his subjects directly and gather, with relative ease, the information that he needs regarding the state of the population. The fishery biologist must resort to indirect census-taking methods if he wishes to determine the stock on hand or to measure fluctuations in abundance in relation to different environmental conditions or changes in fishing regulations.

The abundance of species that are exploited commercially is a matter of considerable economic importance. The accuracy of forecasts of relative abundance is of great concern to fishing interests. The California sardine fishing industry has just experienced an unpredicted and little understood period of scarcity which has caused considerable economic loss.

There is a growing list of publications on the subject of the vital statistics of fish populations and a considerable proportion of the funds for fishery re-

search is spent in this field. Dr. Ricker's work is one of the first attempts to bring together the body of modern knowledge on the subject. He has worked actively in this field himself. His own investigations have included original work on many species of freshwater and anadromous fishes. This first hand experience has been supplemented by a close study of the literature on many marine species, such as the Pacific halibut, the California sardine and many others which are referred to in his publication.

The subject of vital statistics is usually approached by fishery biologists through the data on the catch. After the introductory chapter, which defines the problems and presents certain important numerical relationships and terminology, Dr. Ricker devotes two chapters to the interpretation of catch curves. These curves show the relationship between age and frequency, and provide a means of gaining knowledge of recruitment and mortality. The age of fish can be determined by interpreting the growth markings on the scales or bony structures, or by noting the length frequency groups in the population.

A more direct method of deriving vital statistics is to mark some of the fish that are caught and return them to the water. The marked individuals can then be identified when they are recaptured. Tagging is frequently used, so that individuals can be recognized. The number of marked or tagged fish in the catch is used to calculate the rate of exploitation of the population and to determine the abundance. Dr. Ricker presents the different methods for computing these values and points out the limitations to marking experiments. He also lists the assumptions on which the validity of such population estimates is based. Dr. Ricker devotes six of his nine chapters to the fascinating problems of deriving vital statistics from the results of marking experiments.

He has included a useful "actuarial" appendix. It gives numerical values of the functions needed in this type of work, including the instantaneous mortality rate, the annual mortality rate and the annual survival rate.

Although Dr. Ricker's work will be of most value to fishery "actuaries," others will find it extremely useful as a source of ideas and methods. Military scientists have made use of Volterra's treatment of the problems of competition between species to develop combat models. Dr. Ricker's work may also prove to be useful in the apparently unrelated field of combat attrition.

The work is undoubtedly a first edition. It is hoped that it will be enlarged. In future editions some mention should be made of the approach suggested by Dr. D. B. DeLury and this reviewer, which utilizes the data on the catch per unit of effort to derive direct estimates of the total population under certain conditions.

Elementary Statistical Analysis. *S. S. Wilks* (Professor of Mathematical Statistics, Princeton University, Princeton, N. J.). Princeton, N. J.: Princeton University Press, 1949. Pp. xi, 284. Paper. $2.50. (London E.C.4: Oxford University Press [Amen House, Warwick Square]. 14*s*.)

REVIEWED BY T. A. BANCROFT

Associate Professor, Statistical Laboratory, Iowa State College
Ames, Iowa

CRITERIA must be established for a critical evaluation of the plethora of new textbooks and new editions of old texts on elementary statistics appearing in the postwar period. The usual questions asked in textbook appraisals in general are: Is new material included, i.e., has the author taken cognizance of recent methodological advances and research applications? Again, has the material been arranged in a fresh and stimulating fashion without sacrificing soundness? Final judgment, of course, might ultimately rest on particular questions regarding clarity, simplicity, problems included, references listed, errors, type, general appearance, etc.

It is the reviewer's opinion, however, that, for texts on elementary statistics, a further question is in order, i.e.: What is the particular teaching program in statistics for which the text is designed? That is, was the text designed to be used as a beginning course in statistics in: (*a*) a single field of application such as economics, biology, agriculture, education, etc., or (*b*) in a mathematics department as another course in mathematics or a general service course for applied fields, or (*c*) in a department of statistics as a basic prerequisite course to advanced statistical courses or a general service course for applied fields? Presumably a text designed for a general service course for applied fields might be used either as a mathematics' or statistics' department offering provided mathematical prerequisites were the same. Also, a basic prerequisite course to advanced statistical courses might also use material adequate for a general service course. A companion question is: At what college-year level will the course be offered?

Professor Wilks has attempted with the material and arrangement of this new text to follow the recommendations of three distinguished committees on the teaching of statistics regarding the introduction of a basic elementary course available centrally to all students needing an understanding of statistical concepts and techniques common to all fields of application. Quoting from the preface, Professor Wilks says, "This book has been prepared for a one-semester basic course in elementary statistical analysis which, at Princeton, is the introductory course for all fields of statistical application, and is usually taken in the freshman year. It is especially designed for those who intend to go into the biological and social sciences. It presupposes one semester of elementary mathematical analysis covering topics such as those included in the first half of F. L. Griffin's *Introduction to Mathematical Analysis*."

The text, then, is designed as a basic general service course, presumably to

be offered in a mathematics or statistics department at the latter part of the freshman year. It should be noted that the mathematical prerequisite implies an acquaintance with the elements of calculus which would apparently limit the availability of the text for freshman use to those colleges in which the elements of calculus are taught in the first semester or first two quarters of the freshman year. Within such a college the availability of the text would further be limited for the most part to freshmen whose curricula require the prerequisite mathematical analysis course. In some colleges this would eliminate altogether the students for which the text was primarily prepared, i.e., those in the biological and social sciences. In such cases a possible solution might be to follow still further the recommendations of the three committees mentioned earlier in setting up two basic elementary courses in statistics with different mathematical prerequisites.

A very good indication of the contents of the text may be obtained from the chapter headings: Introduction, Frequency Distribution, Sample Mean and Standard Deviation, Elementary Probability, Probability Distributions, The Binomial Distribution, The Poisson Distribution, The Normal Distribution, Elements of Sampling, Confidence Limits of Population Parameters, Statistical Significance Tests, Testing Randomness in Samples, and Analysis of Pairs of Measurements. No attempt has been made to include a discussion of the analysis of variance or more sophisticated problems in statistical inference since the material is designed for a one semester course only.

In alloting ten chapters to the elements of sampling statistics and only three to descriptive statistics, the reviewer believes that Professor Wilks has made a valiant attempt to provide pertinent material for a modern basic course to be centrally taught in elementary statistics. The text is remarkably free of typographical errors. No list of references is included. Although it is stated that the material is primarily for students in biology and the social sciences, many problems and examples are taken from engineering, industry, and the physical sciences. The reviewer believes that more attention should be given to experimental sampling and that a basic course in statistics should be conducted with benefit of a computing laboratory.

Professor Wilks' experiment in teaching such materials in a central freshman-level course should be observed closely. It represents an approach by a mathematical statistician to provide teaching material for this type course. Attempts to satisfy this need from another approach are being made at other institutions by applied or experimental statisticians. In the reviewer's opinion, the two approaches should supplement rather than compete with one another. Possibly ideal teaching materials for such a course will be the outcome of cooperative efforts of several statisticians with as many viewpoints.

LETTERS ABOUT BOOKS

Readers are invited to submit letters about statistical methodology books for publication in this forum. Concise, informative letters which supplement previously published reviews by pointing out specific strengths, weaknesses, errors, and errata in currently used books are wanted. Criticisms based on actual use of a book as a text are especially desired from statistics instructors. Other letters may consist of suggestions for the writing of books and reviews. Letters which contain adverse criticisms of JOURNAL *reviews will be submitted to the author of the review for any reply he may care to make. Contributors are requested to avoid personalities. The right to decide whether a letter merits publication is reserved. Letters should be sent to the review editor, Oscar K. Buros, Rutgers University, New Brunswick, N. J.*

EXPERIMENTAL DESIGNS IN SOCIOLOGICAL RESEARCH

IN THE literature of social science, two uses of the term "experiment" appear: first, to stand for trial and error attempts at the resolution of some problem of human relations (this is the popular usage); and second, to describe the method of precisely controlled observation used in physical science. It is to be expected that my use of the term, "experimental designs in sociological research," to mean observations of social relations under conditions of control introduced by matching, in my recent book, *Experimental Designs in Sociological Research* (New York: Harper & Bros., 1947), should be misunderstood and hence challenged in reviews by mathematical statisticians Kempthorne,[1] Ackoff,[2] Keyfitz,[3] and Hagood,[4] to mention a few. In fact, at least one sociologist, Hornell Hart,[5] has made substantially the same comment and suggested that the preferred term might be, "statistical comparisons with matched control groups."

The first use of the term "experiment," as a trial and error attempt to influence social relations by social action, I specifically exclude as the meaning of the nine studies summarized in my book (pp. 22–28, 29–33). My interest was to illustrate the crude beginnings of efforts to observe, under conditions of control by matching, what really happened to people when such trial and error experiments, taking the form of programs of social treatment or social reform, were used to influence them. It was my purpose to show that the systematic study of social action (trial and error "experiments"), is necessary if we are to appraise objectively the results often claimed for such programs.

The second meaning, I have also disclaimed (pp. 4–6, 26, 29, 32–33), although the term "experiment" was sometimes used as an abbreviation of the more cumbersome term, "experimental designs in sociological research," but when so used its reference was made clear in the context in which it occurred, even if the original author, whose work was summarized, used the term carelessly (Chap. 4).

These explanations and qualifications still leave us with the need of a term which may be used to describe studies of problems of social relations in which: (1) a group of persons who receive a program of treatment is compared with (2) a group excluded from this treatment; the situation is, (3) such comparisons take place in the natural community (not in the artificial situation of the laboratory or in the class room situation); and finally, (4) embrace an attempt to control by matching some of the factors in the situation other than, (*a*) the pattern

[1] Kempthorne, O. *J. Am. Stat. Assn.*, 43: 489–492, September 1948.
[2] Ackoff, Russell L. *Sci.*, 107: 509–510, May 14, 1948.
[3] Keyfitz, Nathan. *Am. J. Sociol.*, 54: 259–260, November 1948.
[4] Hagood, Margaret Jarman. *J. Am. Stat. Assn.*, 44: 312–313, June 1949.
[5] Hart, Hornell. *Social Forces*, 27: 96–98, October 1948.

of treatment factors, and (*b*) the pattern of response factors. Studies of this sort I have called experimental designs in sociological research. They are not, of course, "experiments" in Fisher's meaning of the term, and I have made no such claim.

The studies which I summarized were conducted in the complex community situation—hence the qualifying terms, "in sociological research." I suspect that the full import of these latter words can hardly be meaningful to any one who has not done empirical research in this field of complex social forces, although several reviewers[6] seem quite aware of their significance. I am still of the opinion that the designation, "experimental designs in sociological research" is a more meaningful descriptive title for these studies than Greenwood's[7] single term "experiment" (that promises too much) which he applied to several of these studies, or Hart's[8] "controlled comparisons" (which promises too little). No doubt usage will determine the acceptance and survival of these terms, as it does all other terms.

Inadequate definitions of the universes sampled is another point of criticism. As to this point, several of the investigators whose work I described did not, unfortunately, offer any such definitions. The chief explanation appears to be that the purpose of each study was to observe what happened to specific groups of subjects when treatment was or was not applied. The studies were exploratory and none was designed to provide a basis for generalization to a universe. Since, however, the criticism is well taken it may be worthwhile to note in passing that there are at least two types of universe concerned where social attitudes are measured. First, there is the universe of respondents in a defined area at a given time. Area sampling technique is the preferred procedure to be followed in all such cases. But two hypothetical universes are also involved and as yet little is known about the sampling techniques to be employed in such cases. One of these universes is that of all possible variations in the attitude of each person on a given issue over a period of time. This universe may be further subdivided into public attitudes and private attitudes, a problem of considerable practical as well as theoretical importance. Then there is the universe of all possible questions that may be formulated to elicit the true attitude of a person on a given issue. Stouffer[9] has recognized this problem by his statement, "Compared with the sampling of respondents, the developments in the sampling of items are still in a relatively primitive state." These points illustrate the great complexity of the concept of universe in sociological research.

The impossibility of randomization of the treated group of persons and the untreated group, presents the most serious problem, and is a well-deserved criticism if randomization has been neglected through ignorance, but randomization was not possible as fully explained in considerable detail (pp. 168–169). To state the problem briefly, no administrator of a private treatment program or of a legalized public reform effort would permit the recipients of treatment to be chosen at random. The local mores of the community or the public law that governs the program determines eligibility for treatment, whether it be by social case work, relief or public housing, by *need* for such treatment. In all such studies of programs we deal with a *de facto* situation which the observer cannot control. It would be an important contribution to research if some way could be found to get around this serious limitation. Furthermore, the consequences of such a limitation both as it precludes generalization from samples to a universe, and as it precludes control of unknown factors, is also discussed in detail (pp. 60, 78, 83, 89–90, 140–141, 166–169, 179–186) so that this point was by no means neglected. It may be also worth noting that because randomization was not possible

[6] See reviews by Hagood (*op. cit.*), R. F. Sletto (*Int. J. Opinion & Att. Res.*, 2: 412–413, Fall 1948), Donald E. Super (*J. Appl. Psychol.*, 33: 93–95, February 1949), T. G. Andrews (*Psychometrika*, 13: 281–283, December 1948), and Otis D. Duncan (*Rural Sociol.*, 13: 199, June 1948).

[7] Greenwood, Ernest. *Experimental Sociology*. New York: King's Crown Press, 1945. Pp. xv, 163.

[8] Hart, op. cit. p. 98.

[9] Stouffer, S. A. "Government and the Measurement of Opinion," p. 436. *Sci. Mo.*, 63: 435–440, December 1946.

no effort was made to use the powerful tool of analysis of variance. This restriction seemed to the author such an obvious logical consequence that no explanation of detailed reasons seemed necessary, although by implication at least, the reviews by mathematical statisticians suggest use of analysis of variance. Perhaps the best statement on this point is still that made by Margaret Hagood[10] in 1941, "Are the methods of analysis of variance (and covariance) applicable for observational situations where data are observed in the cross classifications in which they are found rather than in cross classifications in which they have been randomly placed in an experiment?"

Since the sociologist seems at this stage of development of his field research to be denied the advantages of analysis of variance, resort is made of necessity to the covergence of such evidence as he has: (1) the occurrence of small differences that are in the same direction of logical meaning (pp. 42–46, 49, 103–106); and (2) the persistence of such differences in more and more homogeneous samples after losses from mobility, death, refusals and inability to match have taken their toll. Patterns of difference which still persist after extreme cases have been lost, are regarded as evidence of real differences, and hence may be taken as evidence of the effects of treatment, subject always to the qualifications set by absence of control of unknown factors.

Questions naturally arise as to the validity of applying the conventional tests of statistical significance to differences between the statistics of samples that are non-random. Here again, I can but refer the reader to a somewhat detailed discussion of this point (pp. 176–186) wherein the qualifications and limitations are examined with the conclusion that only empirical results can be claimed, or as one reviewer puts it my solution is, "a pragmatic rather than a fully analytic solution."[11]

My concept of an *ex post facto* experimental design will have to rest on evidence yet to be gathered. Whether it will weather the tests of results in future research only time can tell.

Differences of opinion on the meaning of the concepts "null hypothesis" and "causation" appear in the literature of statistics and of logic. As to the null hypothesis my use follows that of Lindquist,[12] who appears to follow an earlier edition of Fisher's *Design of Experiments*, in which the concept was used to denote any exact hypothesis that one may be interested in disproving, and not merely the hypothesis that a certain parameter is zero, which later has come to be the conventional meaning among some groups of statisticians. Some other statisticians seem also to use the meaning I attach to the term.[13] For the sociologist, whose universes are complex and difficult to define, parameter values are seldom known; but there are real advantages in negative statement of relationships with the object of disproving them, because as I try to explain (pp. 70–73, 83–84, 93, 137–138, 167, 187) such formulations help us get rid of normative considerations. Criticisms of my use of the term seem to be based upon the contention that it departs from the current conventional usage in biological experiments and in biometrics. There are, however, examples of different usages of the same technical concept or term in other areas of research. For example, the term "ambivalence" has one meaning in chemistry and a different but accepted meaning in psychiatry; and the term "rationalization" has one meaning in systematic economics and a different meaning in abnormal psychology.

As to my use of the concept "causation," it will be noted in the first place that no proof of causation is claimed for any of the studies summarized (pp. 50, 74, *passim*), that replication of each study is recommended as the next step toward fuller understanding of the relationships (pp. 31, 57, 90, 120, 139, 176, 177, 185, 188, 189), and that my definition of causation (pp. 52–54) is consistent with the position of logicians as represented in the recent book of Herbert Feigl and Wilfrid Sellers, *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts, Inc., 1949).

F. Stuart Chapin, *Professor of Sociology, Chairman of the Department, and Director of the School of Social Work, University of Minnesota, Minneapolis, Minn.*

[10] *Statistics for Sociologists*, p. 586. New York: Reynal and Hitchcock, Inc., 1941. Pp. viii, 934.

[11] Duncan, *op. cit.*

[12] Lindquist, E. F. *Statistical Analysis in Educational Research*, p. 15. Boston, Mass.: Houghton-Mifflin Co., 1940. Pp. xiii, 266.

[13] Hagood, *op. cit.*, p. 359; and Snedecor, George W. *Statistical Methods*, p. 54. Ames, Iowa: Collegiate Press, Inc., 1937. Pp. xiii, 341.